

Periodic sponge effect on tourism

Hongxuan Yan ¹

*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China and
Center for Forecasting Science, Chinese Academy of Sciences, Beijing 100190, China*

Abstract

Periodic sponge effect on tourism

Keywords: Periodic sponge effect

1 Introduction

Tourism has grown an integral contributor to economics for various countries. Tourist as a goods consumer and a services beneficiary, the huge demands bring the numerous benefits and advantages to the host country. Tourism stimulate the economic growth from different aspects, including directly boosting economic units from the tourism industry (housing, food, transportation, etc), creating new employment opportunities, bringing important revenues to the State budget in the form of taxes and fees and enhancing the developments of other sectors engaging in the accomplishment of the tourism product (Bunghez, 2016). According to the statistical data from the WTO data for Tourism Sector including passenger Transportation Services (excluding freight), Travel Services, and the recreation portion of the Other Commercial Services sector (WTO, 2008), the proportion of the Tourism Sector is the sixth largest sector of the global economy and the largest Service Sector industry in the world (Lew, 2011). Furthermore, the impacts of the development of tourism on the industrial production has been drawn great attentions. Copeland (1991) and Ojaghlou et al. (2019) claimed that a growing tourism will causes the de-industrialization, whereas Kenell (2008) found that there is no evident negative impacts of the development of tourism on the manufacturing industry in Thailand. Besides, tourism and hospitality increase the number of available jobs (Aynalem et al., 2016). Consequently, tourism as a great contributor to economic growth is needed to be investigated to reveal the dynamic mechanism of tourism developments via modelling and forecasting. The characteristic of development paths in tourism area has also been underexplored.

Seasonality as a common characteristic has been widely found in tourism industry (Jolliffe and Farnsworth, 2003). Seasonality in tourism has traditionally been regarded as a major problem which needs to be overcome. Cannas (2012) attempt to provide a rational framework for the tourism seasonality by analysing the main characteristics of these challenges. Butler (1998) studied the characteristics of seasonality and developed a methodology to study this phenomenon. Jang (2004) proposed quantitative solutions via financial portfolio theory to assist marketers in mitigating seasonal effects. Furthermore, in terms of quantitative analysis method, Duro and Turrión-Prats (2019) investigated the causalities of seasonality using a mixed effects panel data model for the main tourist destinations in the world. Ferrante et al. (2018) propose a new index to measure the seasonality in tourism by analysing the pattern of seasonal swing including seasonal amplitude and similarity. This method focus on the ordinal and cyclical structures of seasonal variations. By adopting this approach, a strong relationship between seasonal patterns and the spatial distribution throughout European countries were found. Shen et al. (2009) compared the performance of various econometric time-series models in forecasting seasonal tourism demand and found that the methods of seasonality treatment, such as the pre-test for seasonal unit roots, affect the forecasting performance of the models. Mishra

¹Email: yhx19901122@gmail.com

et al. (2018) evaluated the performance of Holt-Winters and Seasonal ARIMA models for forecasting foreign tourist arrivals in India. Kulendran and Wong (2005) studied ARIMA modelling seasonality in tourism forecasting with two settings, such as one for modeling stochastic nonstationary seasonality and another for a constant seasonality with three seasonal dummies. The out-of-sample forecasting performances were evaluated. Chen et al. (2019) proposed a multiseried structural time series method with one variable to predict seasonal tourism demand. Chang and Liao (2010) adopted seasonal ARIMA model to forecast the monthly outbound tourism departures. Gil-Alana et al. (2008) argued that the simple deterministic model with seasonal dummy variables and AR(1) disturbances has better forecast performance. Saayman and Botha (2017) proposed a SARIMA model with non-linear methods to forecast a seasonal tourist with a structural break in the data. Shen et al. (2011) investigated the performance of combination forecasts in international tourism demand. Chu (2009) forecasted tourism demand with ARMA-based methods.

The long memory phenomena has been widely studied in many areas, which motivates researchers to analyse this non-ignorable dependence between the present observation and all previous observations in a time series (Graves et al., 2014). And the decay rate of this dependence can often be slower than exponential decay. Beran (1994) provided a decent condition for a long memory stationary process using the autocorrelation function (ACF), denoted by $\rho(j)$ for integers j , such that $\sum_{j=-\infty}^{\infty} \rho(j) = \infty$. Furthermore, Granger and Joyeux (1980) and Hosking (1981) proposed autoregressive fractionally integrated moving average (ARFIMA) model by incorporating a fractional differencing operator of certain order d ($0 < d < 1/2$) with the classical autoregressive integrated moving average (ARIMA) model. To capture the real world cyclical phenomena with long-range dependence, Porter-Hudak (1990) and Hassler (1994) adopted the seasonal autoregressive fractionally integrated moving-average (SARFIMA) model. Moreover, to describe the long memory with a oscillatory pattern in many fields, Hosking (1981) extended ARFIMA model to Gegenbauer ARMA (GARMA) model by introducing Gegenbauer polynomial. To overcome the difficulties of applying the classical time series models to model, forecast and analyze time series in the form of counts, Davis et al. (1999) constructed generalised linear ARMA (GLARMA) model by modifying the linear predictor into a ARMA time series structure in a generalised linear regression model. With the prevalence of seasonal discrete time series in many areas such as biology, finance and engineering, Yan et al. (2017) extended GLARMA by replacing the ARMA structure in the linear predictor by Gegenbauer ARMA structure.

1.1 Contribution and structure

From a modelling perspective, our first contribution is to propose generalised linear regression GARMA (GLRGARMA) model and generalised linear regression SARMA (GLRSARMA) model with a innovative function of explanatory variables in order to extend GLGARMA to incorporate relevant information for model fitting and forecast in tourism area. Besides, the generalised Poisson (GP) distribution is adopted to accommodate over- equal- and under-dispersion for certain tourism data. Moreover, the performance of GLRGARMA model and GLRSARMA model with their nested sub-models are compared and evaluated using several well-known selection criteria.

Our second contribution is to investigate the behaviour of tourism data. The pattern of long memory is examined. The analysis of Hurst exponent, ACF plot and periodogram plot shows that Gegenbauer long memory features are presented in tourism data. Furthermore, the distinct characteristics between Gegenbauer long memory and seasonality are demonstrated to reveal the that the GLRGARMA model is more suitable for modelling tourism data.

Our third contribution is to derive a Bayesian approach via the efficient and user-friendly `Rstan` package in estimating our proposed models. For ML approach, the likelihood function is untractable because of involving very high dimensional integrals. Several monitors of convergence of posterior samples are discussed, such as the number of effective sample and \hat{R} estimate. The criteria for modelling performance are also derived.

The rest of the paper is organised as follows. Section 2 reviews GLMs in the insurance area and long

memory time series models for discrete data. The long memory stochastic period effect model is introduced in terms of mean function for eight sub-models and their reduced forms. Section 3 derives the Bayesian approach of our proposed models with various selection criteria. Section 4 conducts a series of simulation studies to validate the accuracy of the estimation techniques we develop for the proposed models. Section 5 presents the in-sample fitting of mortality data, out-of-sample forecast and life table construction. Section 6 concludes the paper.

2 Models

2.1 Long memory time series models with Gaussian innovation

In time series settings, the terms of long memory refers to the strength of statistical dependence, extended temporal dependence or persistence between lagged observations in a time series. And the rate that such lagged dependency decreases should be slower than exponential decay, which is the main feature in the long memory structure time series (Graves et al., 2014). The Wold representation introduced by Wold (1938) states that

Theorem 2.1. *Any zero-mean nondeterministic covariance-stationary process $\mathbf{Y}_{t \in \{1,2,3,\dots,T\}}$ can be expressed as*

$$Y_t = c_t + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \Psi(B)\varepsilon_t + c_t, \quad \varepsilon_t \sim WN(0, \sigma^2), \quad (1)$$

where ε_t and ψ_j are uniquely defined and satisfy $\psi_0 = 1$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, $E(\varepsilon_t \varepsilon_s) = 0, \forall t, s$, the coefficients $\{c_t; t \in \mathbb{Z}\}$ is a deterministic term with $\mathbb{E}(c_t, \varepsilon_s) = 0, \forall t, s$, WN stands for white noise.

Given a stationary time series process $\mathbf{Y}_{1:T} \equiv (Y_1, Y_2, \dots, Y_T)$ which admits Wold representation, with $\mathbf{Y}_{1:T} \in (\mathbb{N} \cup \{0\})^T$, Beran (1994) defined a condition for a long memory stationary process in terms of the divergence of the autocorrelation function (ACF) for Y_t and Y_{t+j} at lag j , such that

$$\lim_{n \rightarrow \infty} \sum_{j=-n}^n |\rho(j)| \rightarrow \infty \quad \text{where} \quad \rho(j) = \frac{\text{Cov}(Y_t, Y_{t+j})}{\sqrt{\text{Var}(Y_t) \text{Var}(Y_{t+j})}}.$$

According to the Szegő-Kolmogorov formula, the spectral density $f_s(\lambda)$ can be derived by taking the Fourier transform of autocovariance functions $\gamma_\theta(l-j) = (\mathbf{\Gamma}_\theta)_{lj}$ and $I_T(\lambda_k)$ can be derived by taking Discrete Fourier Transformation ($\mathfrak{D}(\cdot)$) of $\mathbf{Y}_{t \in \{1,2,3,\dots,T\}}$ where $\lambda_k = \frac{2\pi k}{T}$ for $k = l-j = 1, \dots, [\frac{T}{2}]$, $[\cdot]$ represents the integer part and only half of frequencies are enough to demonstrate the features because of symmetry. The spectral density is given by

$$f_s(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \gamma_\theta(k) \exp(-i\lambda k) dk, \quad -\pi < \lambda < \pi.$$

In practice, the periodogram $I_T(\lambda)$ is usually employed as a estimator of the spectral density $f_s(\lambda)$. Koopmans (1995) stated that $I_T(\lambda)$ is unbiased but inconsistent estimator of $f_s(\lambda_k)$ for a Gaussian white noise

process $\mathbf{Y}_{t \in \{1,2,3,\dots,T\}}$. The periodogram is given by

$$\begin{aligned} I_T(\lambda_k) &= \frac{1}{2\pi T} |\mathfrak{D}(\mathbf{Y}_{t \in \{1,2,\dots,T\}})|^2 \\ &= \frac{1}{2\pi T} \left| \sum_{j=1}^T Y_j e^{-ij\lambda_k} \right|^2 \\ &= \mathbf{Y}_r^2(\lambda_k) + \mathbf{Y}_i^2(\lambda_k), \\ \mathbf{Y}_r(\lambda_k) &:= \frac{1}{\sqrt{2\pi T}} \sum_{j=1}^T \cos(j\lambda_k) Y_j, \\ \mathbf{Y}_i(\lambda_k) &:= \frac{1}{\sqrt{2\pi T}} \sum_{j=1}^T \sin(j\lambda_k) Y_j. \end{aligned}$$

There are two typical types of long memory model structure, the autoregressive fractionally integrated moving average (ARFIMA) model class and the generalised form of ARFIMA called Gegenbauer autoregressive integrated moving average (GARMA) model. Granger and Joyeux (1980) and Hosking (1981) extended the classical ARIMA model to the ARFIMA model, which describe a long memory stationary process with integrated order $d \in (0, 1/2)$. Hosking (1981) further extended to a generalised ARFIMA model called Gegenbauer ARMA (GARMA) model, which can model data with an oscillatory damping autocorrelation function (ACF) because the representation of long memory time series structure of GARMA model is expressed naturally in terms of Gegenbauer polynomials.

For $d \in (0, 1/2)$, the ARFIMA model exhibit long memory features. The short memory ARMA model is a special case of ARFIMA model with $d = 0$ where the long memory operator $(1 - B)^{-2d} = 1$.

Definition 2.1 (ARFIMA). *Consider a stationary time series process with constant $c \in \mathbb{R}$, an ARFIMA model with order (p, d, q) is defined by*

$$\Phi(B)(Y_t - c) = \Theta(B)(1 - B)^{-2d}\varepsilon_t, \quad \varepsilon_t \stackrel{i.i.d}{\sim} N(0, \sigma_\varepsilon^2),$$

where the long memory operator in ARFIMA model can be represented as

$$(1 - B)^{-d_a} = \sum_{j=0}^{\infty} \frac{\Gamma(j + d_a)}{\Gamma(j + 1)\Gamma(d_a)} B^j = \sum_{j=0}^{\infty} \varphi_j B^j \quad \text{with } d_a = 2d.$$

where B is the backshift operator, such that $BY_t = Y_{t-1}$ and

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad \text{and} \quad \Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q,$$

are the autoregressive and moving-average characteristic polynomials, respectively with no common roots.

For $d \in (0, 1/2)$ and $u \in (-1, 1)$, the GARMA model exhibits a long memory with an oscillatory pattern. The ARFIMA model is the speacial case of GARMA model with $u = 1$ (see (Granger and Joyeux, 1980)) such that the factor $(1 - 2uB + B^2)^{-d}$

Definition 2.2 (GARMA). *Consider a stationary time series process with constant $c \in \mathbb{R}$, a GARMA model with order (p, d, q) is defined by*

$$\Phi(B)(Y_t - c) = \Theta(B)(1 - 2uB + B^2)^{-d}\varepsilon_t \equiv \Theta(B) \left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \right), \quad \varepsilon_t \stackrel{i.i.d}{\sim} N(0, \sigma_\varepsilon^2),$$

where $\Phi(B)$ and $\Theta(B)$ are defined in Equations (2.1). The Gegenbauer long memory operator in GARMA model can be represented as

$$(1 - 2uB + B^2)^{-d} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

and ψ_j denote the coefficients of the generating function for the Gegenbauer polynomials $(1 - 2uB + B^2)^{-d}$ (Stein and Weiss, 1971). These coefficients are formulated as

$$\psi_j = \sum_{q=0}^{[j/2]} \frac{(-1)^q (2u)^{j-2q} \Gamma(d - q + j)}{q! (j - 2q)! \Gamma(d)},$$

where $[j/2]$ represents the integral part of $j/2$.

The coefficients ψ_j in Equation (2.2) are functionally dependent on d that controls the strength of long memory and the Gegenbauer parameter u that controls the oscillation of ACF (Rainville, 1960). Furthermore, the coefficients ψ_j can be easily computed using the recursive formula:

$$\psi_j = 2u \left(\frac{d-1}{j} + 1 \right) \psi_{j-1} - \left(2 \frac{d-1}{j} + 1 \right) \psi_{j-2},$$

where the first three terms are $\psi_0 = 1$, $\psi_1 = 2du$ and $\psi_2 = -d + 2d(1 + d)u^2$.

Furthermore, the bounds for the coefficients in the Gegenbauer polynomials ψ_j are the coefficients of ARFIMA φ_j

$$|\psi_j| \leq \frac{(2d)_j}{j!} = \varphi_j,$$

which are demonstrated in Yan et al. (2017).

For a long memory process, investigating the behaviours of ACF with different values of d and u helps to understand the characteristic of long memory time series. In particular, when $u = -1$ and $0 < d < 1/4$, the ACF can be written as

$$\rho(j) = (-1)^j \frac{\Gamma(1 - 2d) \Gamma(j + 2d)}{\Gamma(2d) \Gamma(j - 2d + 1)} \sim \text{constant} \cdot (-1)^j j^{4d-1}, \text{ as } j \rightarrow \infty.$$

For ARFIMA(0, d , 0) with $u = 1$ and $0 < d < 1/4$, it is given asymptotically by,

$$\rho(j) = \frac{\Gamma(1 - 2d) \Gamma(j + 2d)}{\Gamma(2d) \Gamma(j - 2d + 1)} \sim \text{constant} \cdot j^{4d-1}, \text{ as } j \rightarrow \infty.$$

In addition, we note that the closed form ACF for the GARMA(0, d , 0) model with the constraint given by $|u| < 1$, $0 < d < 1/2$ is unavailable (Woodward et al., 1998) but is asymptotically given by

$$\rho(j) \sim \text{constant} \cdot j^{2d-1} \sin(\pi d - j\lambda_0), \text{ as } j \rightarrow \infty,$$

where $\lambda_0 = \cos^{-1}(u)$. For different values of d and u , the ACF plots of long memory time series show different patterns.

2.2 long memory models for time series of counts: GLGARMA

To model discrete time series with Gegenbauer long memory features, Yan et al. (2017) extended the GLM to generalised linear GARMA model (GLGARMA) model which combines the GLM and the GARMA time series model. Hence, the GLGARMA model consisting of an observation variable and state variables

is a generalised state-space model for a time series $\mathbf{Y}_{1:T} \equiv (Y_1, Y_2, \dots, Y_T)$. The canonical log link is adopted as the mean of a count distribution to ensure a strictly positive intensity.

Definition 2.3 (GLGARMA model). *The general model for $\mathbf{Y}_{t \in \{0,1,2,\dots\}}$ of order (p, d, q) is defined as*

$$\begin{aligned} Y_t | \mathcal{F}_{1:t-1}, X_{1:t} &\sim D(\mu_t, \nu), \\ (1 - 2uB + B^2)^d \Phi(B)(\ln \mu_t - c - bX_t) &= \Theta(B)\varepsilon_t, \\ \varepsilon_t &= \begin{cases} z_t \stackrel{i.i.d}{\sim} N(0, \sigma^2) & \text{for PD,} \\ e_t = \frac{Y_{t-1} - \mu_{t-1}}{\sqrt{\mu_{t-1}}} & \text{for OD,} \end{cases} \end{aligned} \quad (2)$$

$D(\cdot, \cdot)$ represents the Poisson, NB, GP or DP distributions. The dispersion parameter is defined as $\nu \in (-1, 1)$ for GP distribution and $\nu > 0$ for NB and DP distributions. $\Phi(B)$ and $\Theta(B)$ are defined in Equations (2.1). The Gegenbauer parameter $|u| < 1$ controls the pattern of oscillation and the long memory parameter $0 < d < 1/2$ determines the strength of long memory.

The data filtration $\mathcal{F}_{1:t-1}$ corresponds to observed realizations of time series up to time $t - 1$. It is the process of choosing a subset of the data set for analysis. So natural sigma algebra $\mathcal{F}_{1:t-1}$ for the observed data filtration means the analysis requires any subset of a given data $(Y_1, Y_2, \dots, Y_{t-1})$ ($Y_t \in (\mathbb{N} \cup \{0\})$) including empty set.

For the error terms in the mean function μ_t , there are two definitions under the parametric driven (PD) and observation driven (OD) modelling approaches and they are defined in equation (2). These state equations combined with the structural equation in (2) form the structure of a class of count valued state space models. The errors of PD model follow a normal distribution which together with the data distributions, Poisson, negative binomial (NB), double Poisson (DP) and generalised Poisson (GP), constitute two sources of randomness whereas the errors of OD model are defined in terms of observations Y_t and mean functions μ_t .

For the data distributions, the Poisson and NB are well-known. GP distribution has the pmf, mean and variance given by

$$\begin{aligned} f(y_t; \mu_t, \nu) &= \mu_t(1 - \nu)[\mu_t(1 - \nu) + \nu y_t]^{y_t-1} e^{-\mu_t(1-\nu)-\nu y_t} / y_t!, \quad \mu_t > 0, \quad -1 \leq \nu < 1, \\ \mathbb{E}(Y_t) &= \mu_t \quad \text{and} \quad \text{Var}(Y_t) = \mu_t(1 - \nu)^{-2}, \end{aligned}$$

respectively. Furthermore, the GP distribution is over-, under- and equi-dispersed when ν is greater than, less than and equal to 0 respectively. On the other hand, the pmf for the DP distribution is given by

$$\tilde{f}(y_t; \mu_t, \nu) = c(\nu, \mu_t) f(y_t; \mu_t, \nu), \quad \mu_t > 0, \quad \nu > 0,$$

where the normalizing term $c(\nu, \mu_t)$, given by Efron (1986), is

$$\frac{1}{c(\nu, \mu_t)} = \sum_{y=0}^{\infty} f(y; \mu_t, \nu) \approx 1 + \frac{1 - \nu}{12\mu_t\nu} \left(1 + \frac{1}{\mu_t\nu} \right).$$

Because of the complicated structure of $c(\nu, \mu_t)$, some properties of the DP distribution are difficult to derive. The unnormalised pmf, mean and variance are given by

$$\begin{aligned} f(y_t; \mu_t, \nu) &= (\nu^{1/2} e^{-\nu\mu_t}) \left(\frac{e^{-y_t} y_t^{y_t}}{y_t!} \right) \left(\frac{e\mu_t}{y_t} \right)^{\nu y_t}, \\ \mathbb{E}(Y_t) &\approx \mu_t \quad \text{and} \quad \text{Var}(Y_t) \approx \frac{\mu_t}{\nu}, \end{aligned}$$

respectively. The DP distribution is over-, under- and equi-dispersed when ν is less than, greater than and equal to 1 respectively.

2.3 Generalised linear regression models with periodic features

The results from Yan et al. (2017) shows that generalised Poisson is consistently the best distribution choice compared with Poisson, NB and DP distributions. Furthermore, PD type models with higher flexibilities is often better than OD type models in both in-sample fitting and put-sample forecast. In terms of model structure, GLGARMA type model outperforms other types model including generalised autoregressive score (GAS) model (Creal et al., 2013), autoregressive conditional Poisson (ACP) model (Groß-KlußMann and Hautsch, 2013) and GLARFIMA model in both short and long memory data with or without obvious seasonal periodicity. In this study, since our models are extended from GLGARMA types model, basing on the conclusions from Yan et al. (2017), only PD type error term and GP distribution with Gegenbauer and seasonal periodic process are considered. Hence, generalised linear regression GARMA (GLRGARMA) model and generalised linear regression SARMA (GLRSARMA) model are proposed to capture the periodic sponge effect in a time series.

Definition 2.4 (GLRGARMA model). *Given a discrete stationary time series process and a natural σ -algebra for the observed data filtration $\mathcal{F}_{1:t-1} = \sigma(Y_1, Y_2, \dots, Y_{t-1})$ with $t \in [1, T]$, a GLGARMA model with order (p, d, q) is defined by*

$$\begin{aligned} Y_t | \mathcal{F}_{1:t-1}, X_{1:t} &\sim GP(\mu_t, \nu), \\ \Phi(B) \ln(\mu_t) &= c + \beta \mathbf{g}(\mathbf{X}_t) + (1 - 2uB + B^2)^{-d} \Theta(B) \varepsilon_t \\ \varepsilon_t &\stackrel{i.i.d}{\sim} N(0, \sigma^2) \end{aligned}$$

where $\mathbf{g}(\mathbf{X}_t) \in \mathbb{R}^W$ are the function of explanatory variables representing the entire feature state at time t with regressors $X_{t,s}$ for $t \in [1, T]$ and $w \in [1, W]$. $\beta \in \mathbb{R}^W$ is a parameter vector that describing the relationship between \mathbf{Y} and \mathbf{X} .

In addition to the model properties and periodic specification, namely, the seasonal difference.

Definition 2.5 (GLRSARMA model). *The GLSARIMA model with order (p, s, q) is given by*

$$\begin{aligned} Y_t | \mathcal{F}_{1:t-1}, X_{1:t} &\sim GP(\mu_t, \nu), \\ \Phi(B)(1 - \alpha B^s) \ln \mu_t &= c + \beta \mathbf{g}(\mathbf{X}_t) + \Theta(B)(1 - \gamma B^s) \varepsilon_t \\ \varepsilon_t &\stackrel{i.i.d}{\sim} N(0, \sigma^2) \end{aligned}$$

where $(1 - B^s)$ is the standard integer seasonal difference operator with the integer seasonal period S that defines the frequency of the seasonal pattern.

2.3.1 ACF oscillation in Gagenbauer fractional differences versus integer and fractional seasonal difference SARFIMA models.

The differences between the SARIMA model and GARMA model are demonstrated by Yan et al. (2017) and clarified that there exists a clear distinct feature between seasonal oscillation and the oscillation that comes from a Gegenbauer long memory process. Their results further showed that the deseasonalisation cannot remove the oscillating behavior in the Gegenbauer long memory time series. The first row in Figure 1 shows the time series plot simulated by SARMA (left panel) and GARMA (right panel). For SARMA model, the period is set to 12 which is agree with our real world data (1 year period for monthly data). For GARMA data set, the long memory parameter is $d = 0.49$ and Gegenbauer parameter is $u = 0.7$, which is similar to the period of SARMA. According to the time series, the tendencies of SARMA and

GARMA looks very similar. It is hardly to distinguish the nature between these two series. The ACF plots (second row) and periodogram (last row) plots are provided to analysis the fundamental characteristics between SARMA and GARMA. The ACF plot for SARMA shows damped periodic peaks with overall short memory pattern. For GARMA, the ACF plot shows a typical Gegenbauer long memory ACF. There exists an oscillated long memory pattern. The periodogram plot as an representation of ACF in frequency domain can easily reveal the differences between SARMA and GARMA. The peaks for SARMA model in periodogram plot are allocated in several places. These peaks must be located in $0, \pi/2$ and π because the number of peaks represents the period, which can be regarded that these peaks chop the region $[0, \pi]$ into 12 pieces. For GARMA model, the location of the peak represents the period of Gegenbauer long memory process, which can be interpret into $\lambda = \cos^{-1}(u)$ where λ is the location of the peak.

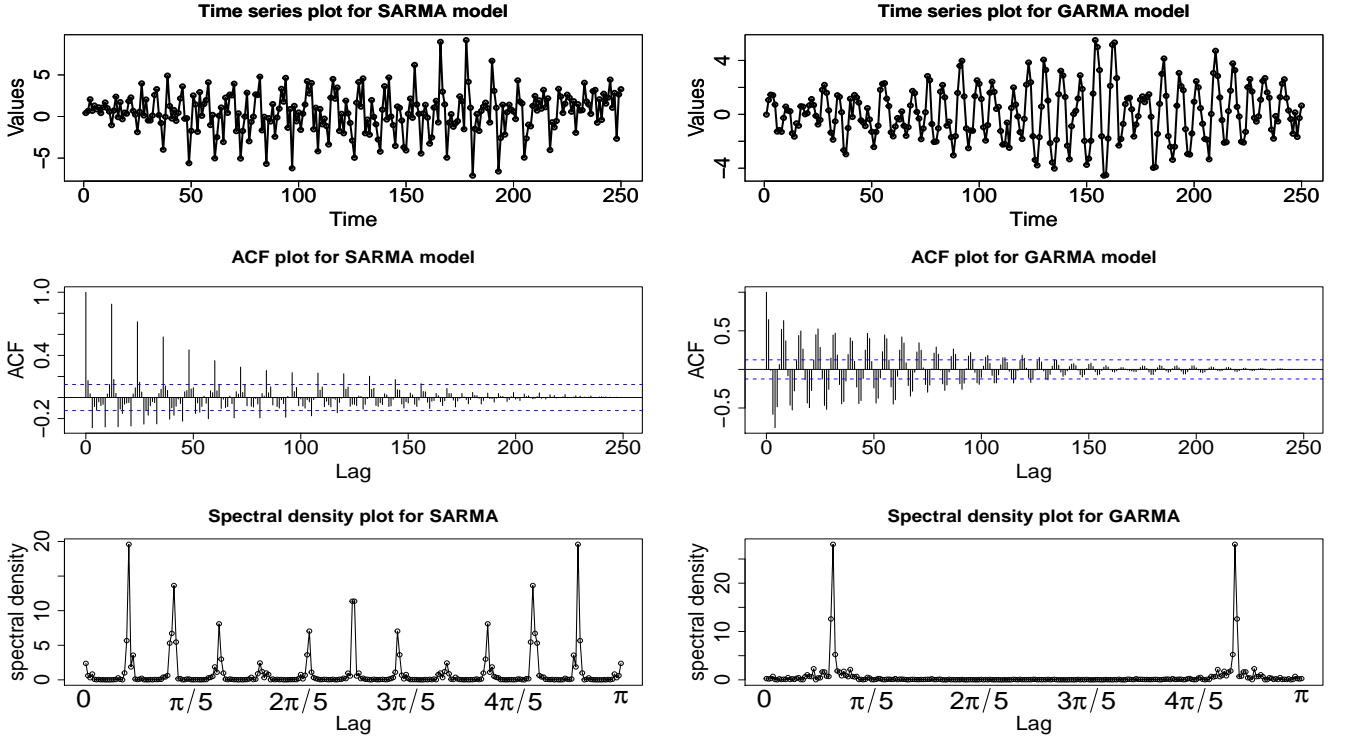


Figure 1: Simulation studies for SARMA and GARMA models

2.3.2 Sub-models.

Considering some special cases of $\beta g(\mathbf{X}_t)$, the GLRGARMA model and GLRSARMA model can be further divided into 8 different sub-models with their mean functions defined in Table 1.

Table 1: 10 different sub-models where model 1,3,5,6 are GLRGARMA models, model 2,7,9,10 are GLR-SARMA models.

Model	mean function
model 1	$\ln \mu_t = \beta_0 + \beta_1 X_1 + (1 - 2uB + B^2)^{-d} \varepsilon_t$
model 2	$\ln \mu_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + (1 - 2uB + B^2)^{-d} \varepsilon_t$
model 3	$\ln \mu_t = (1 - 2uB + B^2)^{-d} X + \varepsilon_t$
model 4	$\ln \mu_t = (1 - 2uB + B^2)^{-d} X + (1 - 2uB + B^2)^{-d} \varepsilon_t$
model 5	$\ln \mu_t = \beta_0 + \beta_1 X_1 + \alpha \ln \mu_{t-s} + (1 - \gamma B^s) \varepsilon_t$
model 6	$\ln \mu_t = \beta_1 X_1 + \beta_2 X_2 + \ln \mu_{t-s} + (1 - B^s) \varepsilon_t$
model 7	$\ln \mu_t = (1 - \delta B^s) X + \alpha \ln \mu_{t-s} + \varepsilon_t$
model 8	$\ln \mu_t = (1 - \delta B^s) X + \alpha \ln \mu_{t-s} + (1 - \gamma B^s) \varepsilon_t$

3 Bayesian inference

In this study, we adopt Bayesian inference that Bayes theorem is used to implement in-sample fitting and out-sample forecast because apriori beliefs on model structures can be incorporated into the state space structures. In addition, the computation complexities in evaluating the marginal likelihood function for PD model which involve latent values and high-dimensional integration can be avoided (Yan et al., 2017). In terms of forecast, posterior predictive distributions can be obtained to provide distributional forecast summaries, which helps to gain a deeper understanding on the model features. For example, the creditable intervals for all parameters can be easily constructed basing on the posterior distributions.

Given a set of observed values $\mathbf{y}_{1:T} = (y_1, y_2, \dots, y_T)$, with each $y_t \in \mathbb{Z}^+ \cup \{0\}$ and the vector of unknown parameters $\boldsymbol{\vartheta}^*$, the posterior distribution for $\boldsymbol{\vartheta}^*$ conditional on $\mathbf{y}_{1:T}$ is given by

$$\pi(\boldsymbol{\vartheta}^* | \mathbf{y}_{1:T}) = \frac{f(\mathbf{y}_{1:T} | \boldsymbol{\vartheta}^*) \pi(\boldsymbol{\vartheta}^*)}{\int f(\mathbf{y}_{1:T} | \boldsymbol{\vartheta}^*) \pi(\boldsymbol{\vartheta}^*) d\boldsymbol{\vartheta}^*} \propto f(\mathbf{y}_{1:T} | \boldsymbol{\vartheta}^*) \pi(\boldsymbol{\vartheta}^*),$$

which is proportional to the likelihood function $f(\mathbf{y}_{1:T} | \boldsymbol{\vartheta}^*)$ and the prior densities $\pi(\boldsymbol{\vartheta}^*)$ estimated using available information or past data. Even if there is no prior information, we can still apply non-informative or reference priors.

3.1 Bayesian model

Let $\boldsymbol{\vartheta}^* = (\boldsymbol{\vartheta}, \mathbf{z})$ denote the vector of all model parameter $\boldsymbol{\vartheta} = (c, u, d, \theta_j, \phi_j, \sigma^2, \nu)$ and state parameter $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_{1:T} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T)$, with each $\varepsilon_t \in \mathbb{R}$. For demonstrating purpose, both GLRGARMA model and GLRSARMA model with simple regression structure βX_t , $p = 0$ for $\Phi(B)$ and $q = 0$ for $\Theta(B)$ are proposed as examples. Hence, for GLRGARMA model, the latent process is $\ln(\mu_t) = c + \beta(X_t) + (1 - 2uB + B^2)^{-d} \varepsilon_t$ and the set of model parameter is $\boldsymbol{\vartheta} = (c, u, d, \beta, \sigma^2, \nu)$.

The priors $\pi(\boldsymbol{\vartheta})$ are defined as:

$$u \sim U(-1, 1), \quad d \sim U(0, 1/2), \quad c \sim N(0, \sigma_c^2), \\ \beta \sim N(0, \sigma_\beta^2), \quad \sigma^2 \sim \Gamma(a, b) \quad \text{and} \quad \nu \sim U(-1, 1),$$

are adopted in which $U(a_u, b_u)$ denotes the uniform priors on the range (a_u, b_u) for the long memory parameters u, d and ν and $\Gamma(a, b)$ denotes the gamma prior with shape and scale parameters a and b respectively for the scale parameter σ^2 . The joint posterior distribution for the GLRGARMA(0, $d, 0$) model with GP

data distribution is

$$\begin{aligned}
f(\boldsymbol{\vartheta}^* | \mathbf{y}_{1:T}, \mathbf{x}_{1:T}) &= f(\mathbf{y}_{1:T} | \boldsymbol{\varepsilon}_{1:T}, \mathbf{x}_{1:T}, \boldsymbol{\vartheta}) f(\boldsymbol{\varepsilon}_{1:T} | \boldsymbol{\vartheta}) \pi(\boldsymbol{\vartheta}) \\
&\propto \prod_{t=1}^T \left[\frac{\exp(c + \beta(X_t) + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j})(1 - \nu) [\exp(c + \beta(X_t) + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j})(1 - \nu) + y_t \nu]^{y_t-1}}{\Gamma(y_t + 1) \exp - y_t \nu - \exp(c + \beta(X_t) + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j})(1 - \nu)} \right] \\
&\quad \times \frac{1}{\sigma} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right) \exp\left(-\frac{c^2}{2\sigma_c^2}\right) \cdot \left(-\frac{\beta^2}{2\sigma_\beta^2}\right) \cdot (\sigma^2)^{a-1} e^{-(\sigma^2)^b} I_u(-1, 1) I_d(0, 0.5) I_\nu(-1, 1),
\end{aligned}$$

where the hyperparameters are set to be $\sigma_c^2 = \sigma_\beta^2 = 10$, $a = 3$ and $b = 1$.

For GLRSARMA model, the latent process is $(1 - B^s) \ln \mu_t = c + \beta \mathbf{g}(\mathbf{X}_t) + (1 - B^s) \varepsilon_t$ and the set of model parameter is $\boldsymbol{\vartheta} = (c, \beta, \sigma^2, \nu)$. The priors $\pi(\boldsymbol{\vartheta})$ are defined as:

$$c \sim N(0, \sigma_c^2), \quad \beta \sim N(0, 1), \quad \gamma \sim N(0, 1), \quad \alpha \sim N(0, 1), \quad \sigma^2 \sim \Gamma(a, b) \quad \text{and} \quad \nu \sim U(-1, 1),$$

The joint posterior distribution for the GLRSARMA(0, s , 0) model with GP data distribution is

$$\begin{aligned}
f(\boldsymbol{\vartheta}^* | \mathbf{y}_{1:T}, \mathbf{x}_{1:T}) &= f(\mathbf{y}_{1:T} | \boldsymbol{\varepsilon}_{1:T}, \mathbf{x}_{1:T}, \boldsymbol{\vartheta}) f(\boldsymbol{\varepsilon}_{1:T} | \boldsymbol{\vartheta}) \pi(\boldsymbol{\vartheta}) \\
&\propto \prod_{t=1}^T \left[\frac{\exp(c + \beta(X_t) + \alpha \ln \mu_{t-s} + (1 - \gamma B^s) \varepsilon_{t-j})(1 - \nu)}{\Gamma(y_t + 1) \exp - y_t \nu - \exp(c + \beta(X_t) + \alpha \ln \mu_{t-s} + (1 - \gamma B^s) \varepsilon_{t-j})(1 - \nu)} \right. \\
&\quad \times [\exp(c + \beta(X_t) + \alpha \ln \mu_{t-s} + (1 - \gamma B^s) \varepsilon_{t-j})(1 - \nu) + y_t \nu]^{y_t-1} \left. \right] \\
&\quad \times \frac{1}{\sigma} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right) \exp\left(-\frac{c^2}{2\sigma_c^2}\right) \cdot \left(-\frac{\beta^2}{2\sigma_\beta^2}\right) \cdot (\sigma^2)^{a-1} e^{-(\sigma^2)^b} I_\nu(-1, 1).
\end{aligned}$$

3.2 Bayesian forecasting

One outstanding advantage of Bayesian inference in forecasting is the construction of posterior predictive distributions for all forecasts. In this study, m -step forecasts $\mathbf{y}_{T+1:T+m}$ are constructed via a sequence of 1-step ahead forecasts of y_{T+s} , $s = 1, \dots, m$ using sliding window where the observed data filtration for each window is $\mathcal{F}_{s:T+s-1}$ and $\mathbf{x}_{1:T}$. The posterior predictive distribution for y_{T+s} , $s = 1, \dots, m$ is defined as

$$\begin{aligned}
&f(y_{T+s} | \mathcal{F}_{s:T+s-1}, \mathbf{x}_{s:T+s}) \\
&= \int \dots \int f(y_{T+s} | \mu_{T+s}, \boldsymbol{\vartheta}, \mathcal{F}_{s:T+s-1}, \mathbf{x}_{s:T+s}) f(\mu_{T+s} | \boldsymbol{\mu}_{s:T+s-1}, \boldsymbol{\vartheta}, \mathcal{F}_{s:T+s-1}, \mathbf{x}_{s:T+s}) f(\boldsymbol{\vartheta} | \mathcal{F}_{s:T+s-1}, \mathbf{x}_{s:T+s}) d\boldsymbol{\mu}_{s:T+s} d\boldsymbol{\vartheta},
\end{aligned}$$

and this integral can be approximated by the Monte Carlo estimator, constructed from posterior samples according to:

$$\hat{f}(y_{T+s} | \mathcal{F}_{s:T+s-1}, \mathbf{x}_{s:T+s}) = \frac{1}{L} \sum_{l=1}^L f(y_{T+s} | \boldsymbol{\mu}_{s:T+s}^{(l)}, \boldsymbol{\vartheta}_s^{(l)}, \mathcal{F}_{s:T+s-1}, \mathbf{x}_{s:T+s}).$$

In this study, we set $L = 90,000$ as the number of iterations after burn-in in each MCMC sampler run given the current window of information $\mathcal{F}_{s:T+s-1}$ and $\mathbf{x}_{s:T+s}$. In addition, $\boldsymbol{\mu}_{s:T+s}^{(l)}$ and $\boldsymbol{\vartheta}_s^{(l)}$ are the l -th draw in the posterior sample of $\boldsymbol{\mu}_{s:T+s}$ and $\boldsymbol{\vartheta}$, respectively. For Bayesian inference, except of the posterior predictive distribution, the various posterior predictive point estimators and predictive credible intervals can also be obtained. Empirical Bayes forecasts as a typical forecast estimator in the Bayesian setting can improve the computational efficiency (?). Moreover, the features of frequentist and empirical Bayes forecasts are

compared by ?. For empirical Bayes forecasts, the calculations undertake condition upon selected posterior (in sample) point estimators denoted by $\tilde{\boldsymbol{\vartheta}}$ and $\tilde{\boldsymbol{\mu}}_{s:T+s}$ rather than integrating out posterior (in sample) parameter uncertainty from the predictive distribution and resultant forecast estimators (Yan et al., 2017).

$$\hat{f}^{EB}(y_{T+s}|\mathcal{F}_{s:T+s-1}) = f(y_{T+s}|\tilde{\boldsymbol{\mu}}_{s:T+s}, \tilde{\boldsymbol{\vartheta}}_s, \mathcal{F}_{s:T+s-1}).$$

Typically the point estimators used in $\tilde{\boldsymbol{\vartheta}}_s$ (similarly for $\tilde{\boldsymbol{\mu}}_{s:T+s}$) are either formed from the maximum-a-posteriori estimate (MAP) or the estimate which minimizes the posterior expected loss. The concept of MAP is similar to the ML estimate when the priors are uninformative since in this case $\tilde{\boldsymbol{\vartheta}}_s$ is the mode of the posterior distribution

$$\tilde{\boldsymbol{\vartheta}}_{s,MAP} = \arg \max_{\boldsymbol{\vartheta}_s} f_{\boldsymbol{\vartheta}_s}(\boldsymbol{\vartheta}_s|\mathcal{F}_{s:T+s-1}).$$

Another Bayes estimator which minimises the posterior expected loss (PEL) is defined as

$$\tilde{\boldsymbol{\vartheta}}_{s,PEL} = \arg \min_{\boldsymbol{\vartheta}_s} E(L(\boldsymbol{\vartheta}_s, \tilde{\boldsymbol{\vartheta}}_s|\mathcal{F}_{s:T+s-1})),$$

where $L(\boldsymbol{\vartheta}_s, \tilde{\boldsymbol{\vartheta}}_s|\mathcal{F}_{s:T+s-1})$ is the loss function. One example is the commonly used minimum mean square error (MSE) estimator defined as

$$\tilde{\boldsymbol{\vartheta}}_{s,MSE} = \arg \min_{\boldsymbol{\vartheta}_s} E([\boldsymbol{\vartheta}_s - \tilde{\boldsymbol{\vartheta}}_s]^2|\mathcal{F}_{s:T+s-1}),$$

where $\tilde{\boldsymbol{\vartheta}}_{s,MSE}$ corresponds to the posterior mean $E(\boldsymbol{\vartheta}_s|\mathcal{F}_{s:T+s-1}) = \bar{\boldsymbol{\vartheta}}_s$. If the minimum absolute error (AE) estimator $L(\boldsymbol{\vartheta}_s, \tilde{\boldsymbol{\vartheta}}_s) = |\boldsymbol{\vartheta}_s - \tilde{\boldsymbol{\vartheta}}_s|$ is used, it gives $\tilde{\boldsymbol{\vartheta}}_{s,AE} = \boldsymbol{\vartheta}_{s,0.5}$ which is the posterior median.

3.3 Bayesian tool: implementation with Rstan

To implement the proposed models efficiently, we choose the Bayesian R package `Rstan` which utilises the `stan` program within R developed in the C++ language. The sampling method applied in `Rstan` is the Hamiltonian Monte Carlo (HMC) sampler (see (Duane et al., 1987), (Neal, 1994)) that belongs to the class of Markov chain Monte Carlo (MCMC) sampling methods (see (Metropolis et al., 1953)). For some complicated models with many parameters, the HMC sampler converges faster than the conventional samplers such as random-walk Metropolis (see (Metropolis et al., 1953)) and Gibbs sampler (see (Geman and Geman, 1984)) because the HMC sampler transforms the simulation of high-dimensional target distribution into the simulation of Hamiltonian dynamics (see (Neal, 2011)). By using the gradient information, Hamiltonian dynamics can produce distant proposals for the Metropolis algorithm, thereby avoiding the slow exploration of the state space that results from the diffusive behaviour of random-walk proposals.

For the HMC sampler, to assess the dependence, precision and convergence of posterior sample, three measures are reported in `Rstan`. The first measure is the *number of effective samples* which indicates dependence within a Monte Carlo sample. The second measure is the *Monte Carlo standard error* (MCSE)

$$\text{MCSE} = \frac{\text{posterior standard deviation}}{\sqrt{\text{number of effective samples}}},$$

which reports the error of estimation for the posterior mean. To monitor the convergence for $k > 2$ chains of length $2n$ each, Gelman and Rubin (1992) proposed \hat{R} which is defined as

$$\hat{R} = \frac{\hat{V}}{W} \cdot \frac{df}{df - 2},$$

where

$$\hat{V} = \frac{n-1}{n}W + \frac{k+1}{kn}B, \quad W = \sum_{i=1}^k \frac{s_i^2}{k}, \quad B = n \sum_{i=1}^k \frac{(\bar{\vartheta}_i - \bar{\vartheta}_{..})^2}{k-1}, \quad df = \frac{2\hat{V}^2}{\hat{\text{Var}}(\hat{V})}, \quad (3)$$

$$\begin{aligned} \hat{\text{Var}}(\hat{V}) = & \left(\frac{n-1}{n} \right)^2 \frac{1}{k} \hat{\text{Var}}(s_i^2) + \left(\frac{k+1}{kn} \right)^2 \frac{2}{k-1} B^2 + \\ & 2 \frac{(k+1)(n-1)}{kn^2} \cdot \frac{n}{k} [\hat{\text{Cov}}(s_i^2, \bar{\vartheta}_{i..}) - 2\bar{\vartheta}_{..} \hat{\text{Cov}}(s_i^2, \bar{\vartheta}_{i..})], \end{aligned} \quad (4)$$

s_i is the within-chain variance and ϑ_{ij} is the j -th parameter in chain i . If \hat{R} is close to 1, the parameter ϑ has converged.

In this study, the number of chain is $k = 1$ and hence $B = 0$ in (3) and (4). For this chain, there are overall totally 100,000 iterations, with the first 10,000 iterations as burn-in. Hence, there are $L = 90,000$ subsequent iterations with thin set to 1. The values of \hat{R} for each estimators and the history plot are carefully checked to ensure that all parameters meet the convergence condition. For all of in-sample fitting and out sample forecast studies, the number of effective sample ranges from 75,000 to 86,000 across parameters. The range of \hat{R} is between 1.0000 and 1.0003, which indicates moderate dependency and clear convergence.

3.4 Model selection and forecast performance

The performance of each model is evaluated through a popular Bayesian model selection criteria called deviance information criterion (DIC) (Spiegelhalter et al., 2014). As a generalisation of Akaike's Information Criterion (AIC), DIC can deal with models containing informative priors, such as hierarchical models. As the priors can effectively restrict the freedom of model parameters, the number of parameters as required in the calculation of AIC is generally unclear. DIC overcomes such problems by providing an estimate for the effective number of parameters. The DIC can be calculated using the equation

$$\text{DIC} = \bar{D} + p_D = 2\bar{D} - D(\bar{\vartheta}_x),$$

where the deviance is defined as $D(\vartheta_x) = -2 \ln(f(\mathbf{y}_x | \vartheta_x))$, $\bar{D} = \mathbb{E}_{\vartheta_x | y_x} [-2 \ln(f(\mathbf{y}_x | \vartheta_x))]$ measures the model fit and the estimated number of parameters $p_D = \bar{D} - D(\bar{\vartheta}_x)$ measures model complexity (Spiegelhalter et al., 2002).

Consider the m -step ahead forecasts $\hat{y}_{x,t}$ given by the posterior mean or median and the observations $y_{x,t}$ with T time point and g age groups, the forecast performance can be evaluated by adopting three types of measures, namely residuals $r_{x,t} = y_{x,t} - \hat{y}_{x,t}$, percentage errors $p_{x,t} = \frac{r_{x,t}}{y_{x,t}} \times 100$ and scaled errors $\epsilon_{x,t}$ defined in Equation (7). Based on $r_{x,t}$ and $p_{x,t}$, three popular criteria, namely mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE), are defined respectively below

$$\begin{aligned} \text{MAE} &= \frac{1}{g} \sum_{x=1}^g \left[\frac{1}{m} \sum_{t=1}^m |r_{x,T+t}| \right], \quad \text{RMSE} = \sqrt{\frac{1}{g} \sum_{x=1}^g \left[\frac{1}{m} \sum_{t=1}^m r_{x,T+t}^2 \right]} \\ \text{and } \text{MAPE} &= \frac{1}{g} \sum_{x=1}^g \left[\frac{1}{m} \sum_{t=1}^m |p_{x,T+t}| \right]. \end{aligned} \quad (5)$$

However $r_{x,t}$ are scale-dependent making comparison difficult. Although $p_{x,t}$ are scale-free, they are sensitive to observations close to zero. Hence the fourth criterion we adopt is mean absolute scaled error (MASE)

defined as

$$\text{MASE} = \frac{1}{g} \sum_{x=1}^g \left[\frac{1}{m} \sum_{t=1}^m |\epsilon_{x,T+t}| \right], \quad (6)$$

making use of the scaled errors

$$\epsilon_{x,T+t} = \frac{r_{x,T+t}}{\frac{1}{m-1} \sum_{t=2}^m |y_{x,T+t} - y_{x,T+t-1}|}, \quad (7)$$

proposed by Hyndman and Koehler (2006). Furthermore, similar approach can also be applied to evaluate estimated results $\hat{\mu}_{x,t}$ calculated by the posterior mean or median. Hence, the residuals $r_{x,t}^s = \mu_{x,t} - \hat{\mu}_{x,t}$, percentage errors $p_{x,t}^s = \frac{r_{x,t}^s}{\mu_{x,t}} \times 100$ and scaled errors

$$\epsilon_{x,t}^s = \frac{r_{x,t}^s}{\frac{1}{m-1} \sum_{t=2}^m |\mu_{x,t} - \mu_{x,t-1}|},$$

can be used to construct similar criteria for the μ estimator, namely the MAE, RMSE, MAPE and MASE by using the same formulas in Equations (5) to (6).

4 Data analysis

Denmark is often be analysed in political-economic research area because of the lowest Gini coefficient (latest OECD figures from 2012), strong local companies with great competitiveness and prominent economic performance (Refslund and Sørensen, 2016). The data set analysed in this study is obtained from Den (2020) that is the central authority on Danish statistics. It is a state institution under the Ministry of Economic Affairs and the Interior. They collect, compile and publish statistics on the Danish society. Danish was a predominantly agricultural country. After the year of 1945, Denmark has significantly developed the industrial base and service sector. By 2017, the agriculture sector only contribute less than 2% of overall GDP. On the other hand, the industrial base and services contribute around 18% and 76%, respectively (Den, 2020). And tourism is the most important component in Denmark's tertiary industry, which can be a representative indicator of service sector. In this study, we adopt rented hotel room numbers in a month as a index to investigate the dynamic mechanism of service sector statistics. Furthermore, we use industrial production index (IPI) to describe the changes in the manufacturing because the IPI is a key indicator to measure the output of industrial economic activities. In order to describe the characteristic of agriculture in Denmark, we adopt the percentage of agriculture in Gross Domestic Product (GDP) as an index. Agriculture sector includes forestry, hunting, and fishing, cultivation of crops and livestock production. Besides, The labour market of Denmark is the freest in Europe (Wor, 2020). Employers maintain a very high level of flexibility, which means they can hire and fire whenever they want. And on the other hand, the unemployment compensation is relatively high to guarantee a stable living standard for unemployed persons. The unemployed rate used in this paper is the statistic for unemployment comprise all unemployed persons basing on the resident population in Denmark. In addition, the statistics of electricity production is used to measure the industrial production activities.

4.1 Empirical data analysis

The Hurst exponent H also known as the index of long-range dependence was proposed by Hurst (1951). It is a classical self-similarity parameter that measures the long memory feature in a time series (Millen and Beard, 2003). Since it is robust with few assumptions about underlying system, it has been widely

applied to many fields (Qian and Rasheed, 2004). A value of H in the range $(\frac{1}{2}, 1)$ indicates long memory in a time series, which means that a high value in the series will more likely be followed by another high value and such an effect is likely to maintain for a long period into the future. A value of $H = \frac{1}{2}$ can indicate a standard Brownian motion which is a short memory process. Furthermore, There exists a relationship between d and H , which is then given by $d = H - 0.5$. Consequently, the estimator of the Hurst exponent H can be approximate the long memory parameter d (Yan et al., 2017). There exist various estimators of H , in this section, a well-known estimator called rescaled range analysis (R/S) is adopted to implement following empirical studies.

Hurst (1951) proposed the first Hurst exponent estimator using the rescaled range R/S analysis to measure the intensity of long-range dependence. Given a time series $Y_{t \in \{1,2,3,\dots,T\}}$, the sample mean and the standard deviation process are given by

$$\bar{Y}_T = \frac{1}{T} \sum_{j=1}^T Y_j \quad \text{and} \quad S_t = \sqrt{\frac{1}{t-1} \sum_{j=1}^t (X_j)^2}, \quad (8)$$

where the mean adjusted series $X_t = Y_t - \bar{Y}_T$. Then a cumulative sum series is given by $Z_t = \sum_{j=1}^t X_j$ and the cumulative range based on these sums is

$$R_t = \text{Max}(0, Z_1, \dots, Z_t) - \text{Min}(0, Z_1, \dots, Z_t). \quad (9)$$

An important proposition for the estimator of H was derived by Mandelbrot (1975).

Proposition 4.1. Consider a time series $Y_t \in \mathbb{R}$ and define S_t and R_t in Equations (8) and (9) respectively, then $\exists C \in \mathbb{R}$ such that the following asymptotic property of the rescaled range R/S holds

$$[R/S](T) = \frac{1}{T} \sum_{t=1}^T R_t/S_t \sim CT^H, \quad \text{as } T \rightarrow \infty.$$

In addition, for small sample size T , the rescaled range R/S can also be approximated by following equation (Annis and Lloyd, 1976)

$$[R/S](T) = \begin{cases} \frac{T-1/2}{T} \frac{\Gamma((T-1)/2)}{\sqrt{\pi(T/2)}} \sum_{j=1}^{T-1} \sqrt{\frac{T-j}{j}}, & \text{for } T \leq 340 \\ \frac{T-1/2}{T} \frac{1}{\sqrt{T\pi/2}} \sum_{j=1}^{T-1} \sqrt{\frac{T-j}{j}}, & \text{for } T > 340 \end{cases}$$

where the $\frac{T-1/2}{T}$ term was added by Peters (1994). The H estimate can be obtained by a simple linear regression

$$\log R/S(T) = \log C + H \log T.$$

Hence, the definition for the estimator of H is given by the following equation

Definition 4.1 (Estimator \hat{H} by R/S). The estimator \hat{H} based on the rescaled range R/S analysis is given by

$$\hat{H}_{R/S} = \frac{T(\sum_{t=1}^T \log R/S(t) \log t) - (\sum_{t=1}^T \log R/S(t))(\sum_{t=1}^T \log t)}{T(\sum_{t=1}^T (\log t)^2) - (\sum_{t=1}^T \log t)^2}. \quad (10)$$

The empirical confidence interval of \hat{H} given in Equation (10) with sample size $T = 2^N$ (Weron, 2002) is

$$(0.5 - \exp(-7.33 \log(\log N) + 4.21), \exp(-7.20 \log(\log N) + 4.04) + 0.5).$$

In this paper, the R package called `pracma` is adopted to estimate the value of H adopting the R/S analysis. For the data set of rented hotel room numbers in a month, the estimated Corrected R over S Hurst exponent is

0.908, which indicates that there exist a strong long memory in this data. Furthermore, according to Figure 2 that shows the ACF and periodogram plot for rented hotel room numbers, the class of long memory structure is a typical Gegenbauer long memory pattern with apparent oscillatory structure in ACF plot. For the periodogram plot, the peaks locates at non-zero position which aligns with the characteristics for Gegenbauer long memory type models.

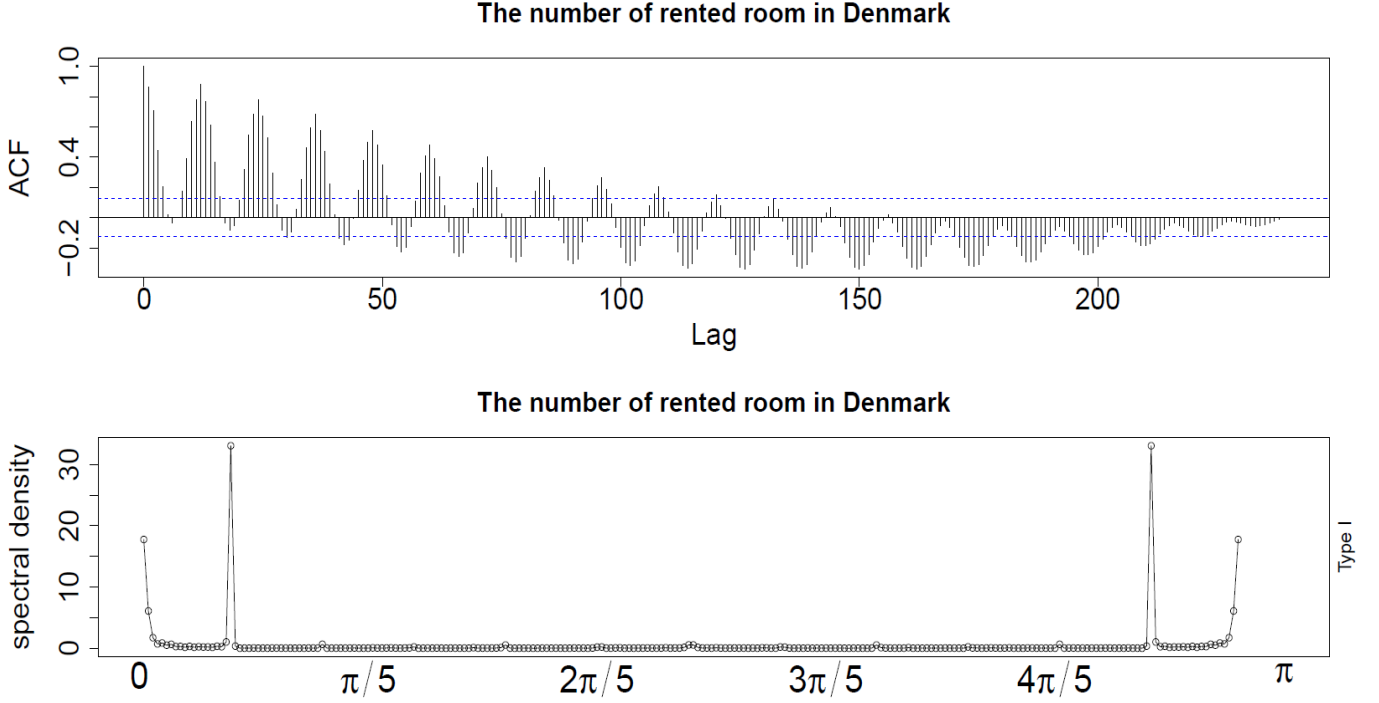


Figure 2: Plot of ACF and periodogram for rented hotel room numbers

4.2 Model fitting

In this section, Model 1 to Model 8 are adopted to fit rented hotel room numbers in a month. Industrial production index (IPI), power production (PG) and unemployment rate (UR) are incorporated to improve model feasibility. The in-sample fitting performances of seasonal component and Gegenbauer long memory component are compared in this study. To monitoring the convergence of Bayesian approach, the values of \hat{R} for each estimators are between 1.0000 and 1.0003 and the number of effective sample are always more than 80,000. Figure 3 is an example of convergence test, which reports the MCMC sample path for several key parameters for Model 1. According to these plots, the model parameters are properly estimated.

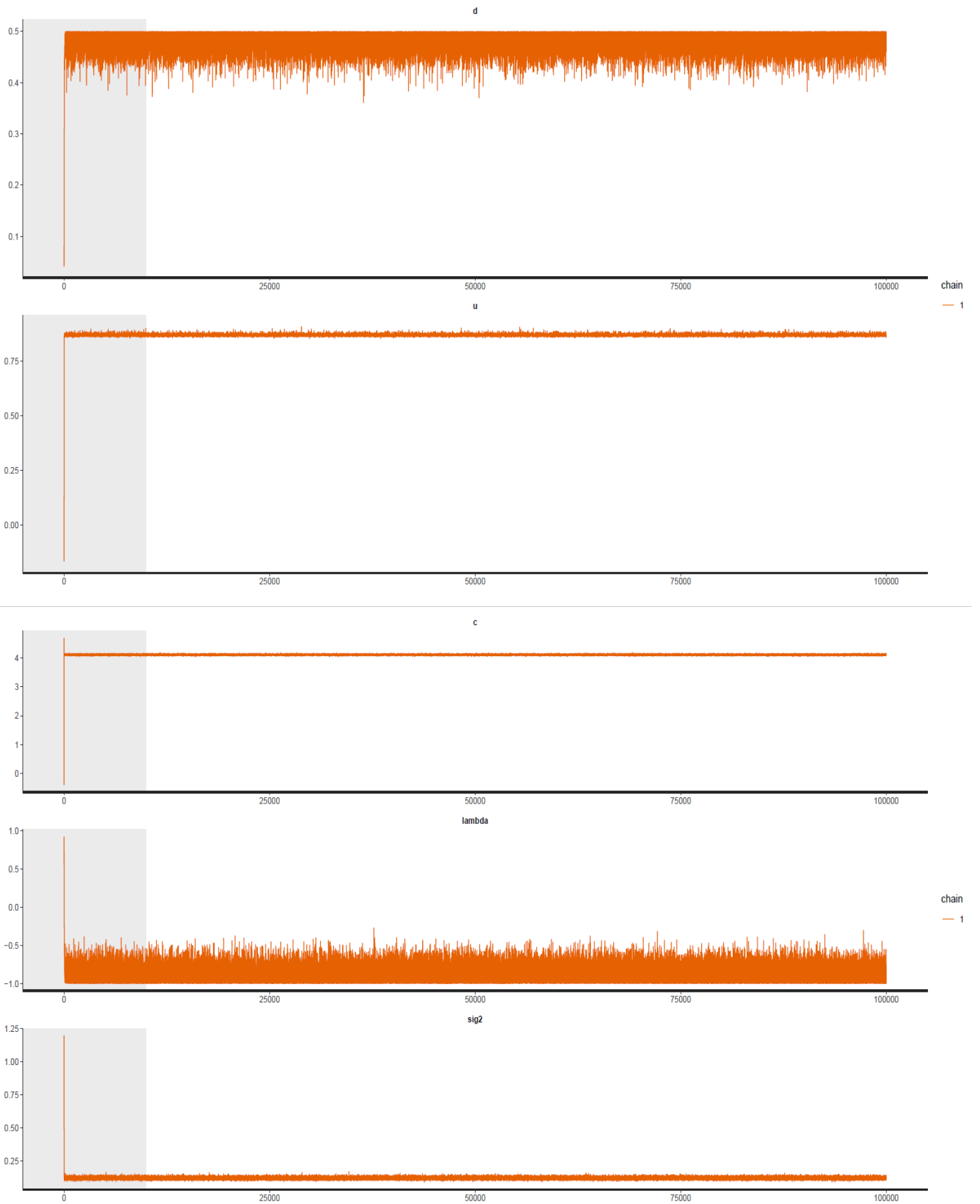


Figure 3: Trace plots for part of Model 1 parameters

The goodness of fit for the eight models are evaluated by using DIC to select the best fitting model for each data set. Table 2 reported the DIC values of these models. The performance of the models incorpo-

rating Gegenbauer long memory or seasonal component into error terms are significantly better than the models with these components on the explanatory variables X . This indicates that the historical information provided by the explanatory variables X will cause negative impacts on modelling performance. As a contrast, the error terms only keep essential characteristics of previous knowledge for modelling. Furthermore, the performance of model with seasonal component is similar with the long memory model, since the differences of DIC values between both type models are very small. Consequently, for in-sample fitting study, Model 1, Model 2, Model 5 and Model 6 outperform other models.

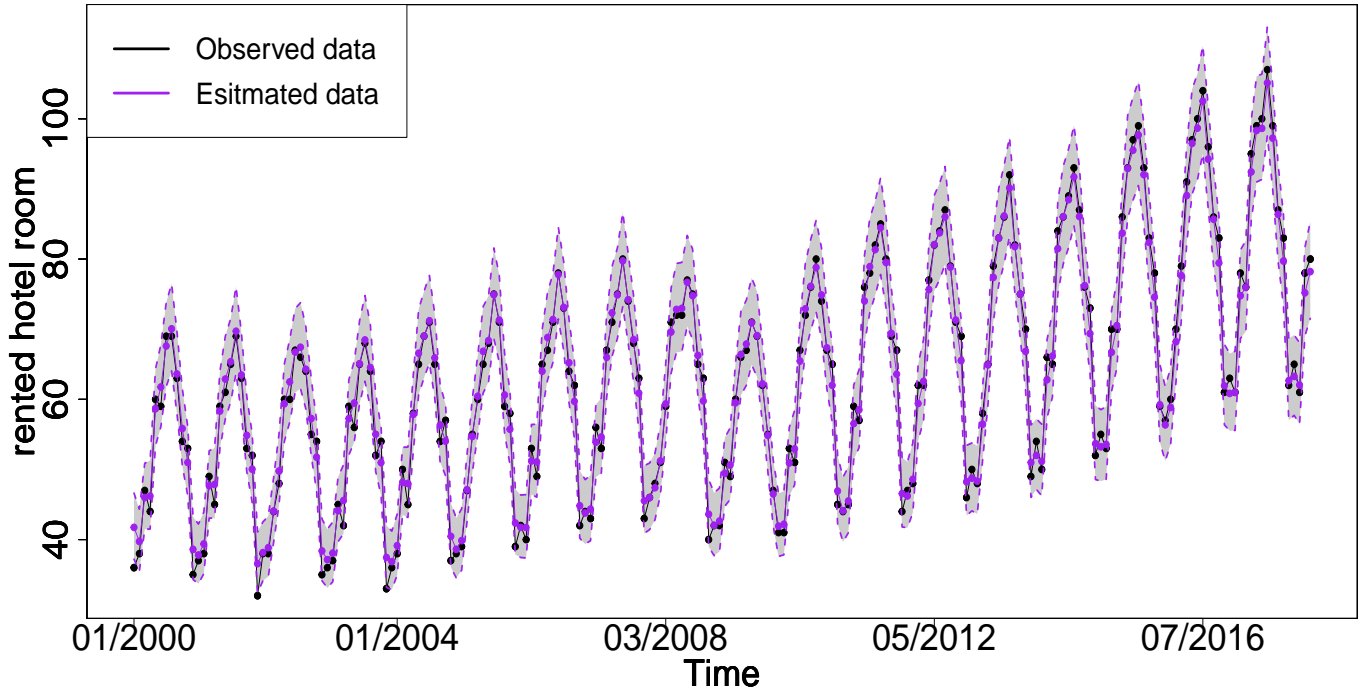
Table 2: DIC results.

Model 1 DIC	$X = \text{IPI}$ 1285.48	$X = \text{UR}$ 825.40	$X = \text{PG}$ 1440.54
Model 2 DIC	$X_1 = \text{IPI}, X_2 = \text{PG}$ 1325.06	$X_1 = \text{IPI}, X_2 = \text{UR}$ 794.77	$X_1 = \text{UR}, X_2 = \text{PG}$ 863.48
Model 3 DIC	$X = \text{IPI}$ 2367.663	$X = \text{UR}$ 2115.049	$X = \text{PG}$ 2510.72
Model 4 DIC	$X = \text{IPI}$ 2416.19	$X = \text{UR}$ 2291.37	$X = \text{PG}$ 2516.81
Model 5 DIC	$X = \text{IPI}$ 1287.67	$X = \text{UR}$ 825.87	$X = \text{PG}$ 1327.78
Model 6 DIC	$X_1 = \text{IPI}, X_2 = \text{PG}$ 1318.26	$X_1 = \text{IPI}, X_2 = \text{UR}$ 820.23	$X_1 = \text{UR}, X_2 = \text{PG}$ 847.93
Model 7 DIC	$X = \text{IPI}$ 2397.63	$X = \text{UR}$ 2187.52	$X = \text{PG}$ 2508.97
Model 8 DIC	$X = \text{IPI}$ 2417.82	$X = \text{UR}$ 2128.53	$X = \text{PG}$ 2607.18

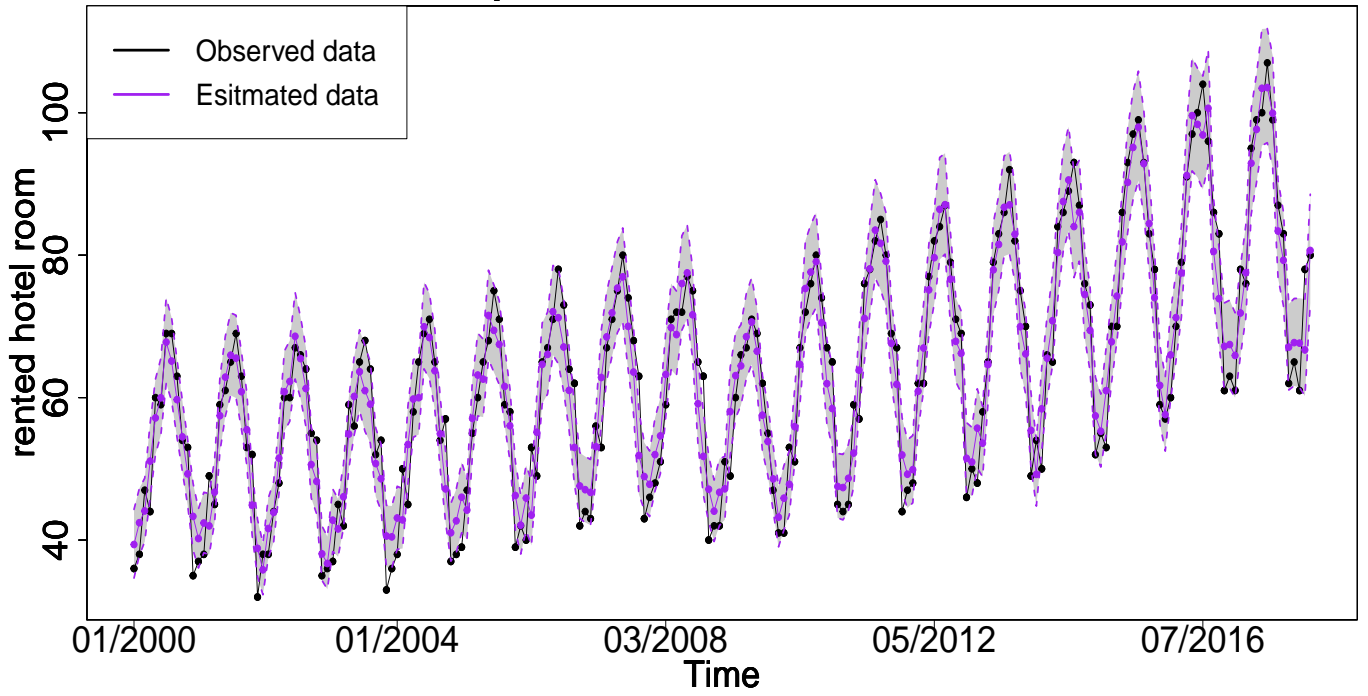
where HR is short for the number of rented hotel room, PG is short for power production, and UR is short for unemployment rate.

The figures show the in-sample fit performance, which confirms that there is no significant evidence to claim long memory model superior than seasonal model in model fitting.

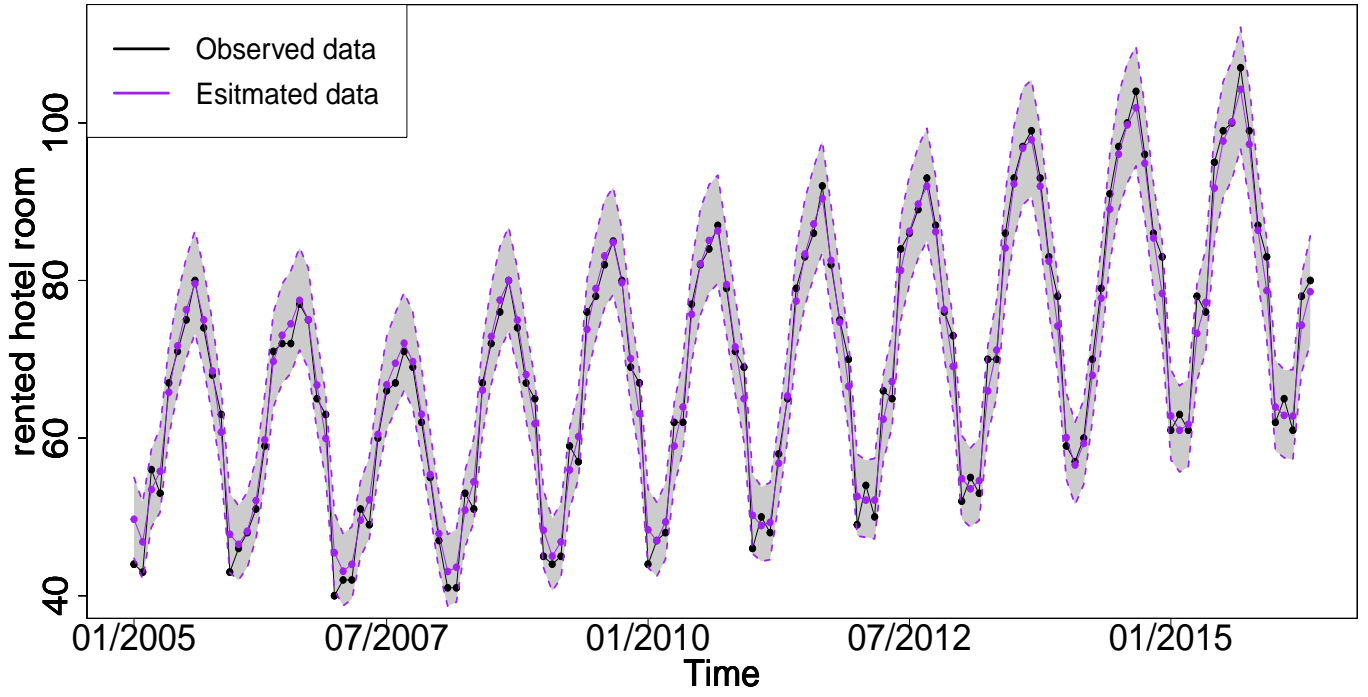
In-sample fit results of GLRGARMA model

Figure 4: In-sample fitting plot for Model 1 with $X = \text{IPI}$

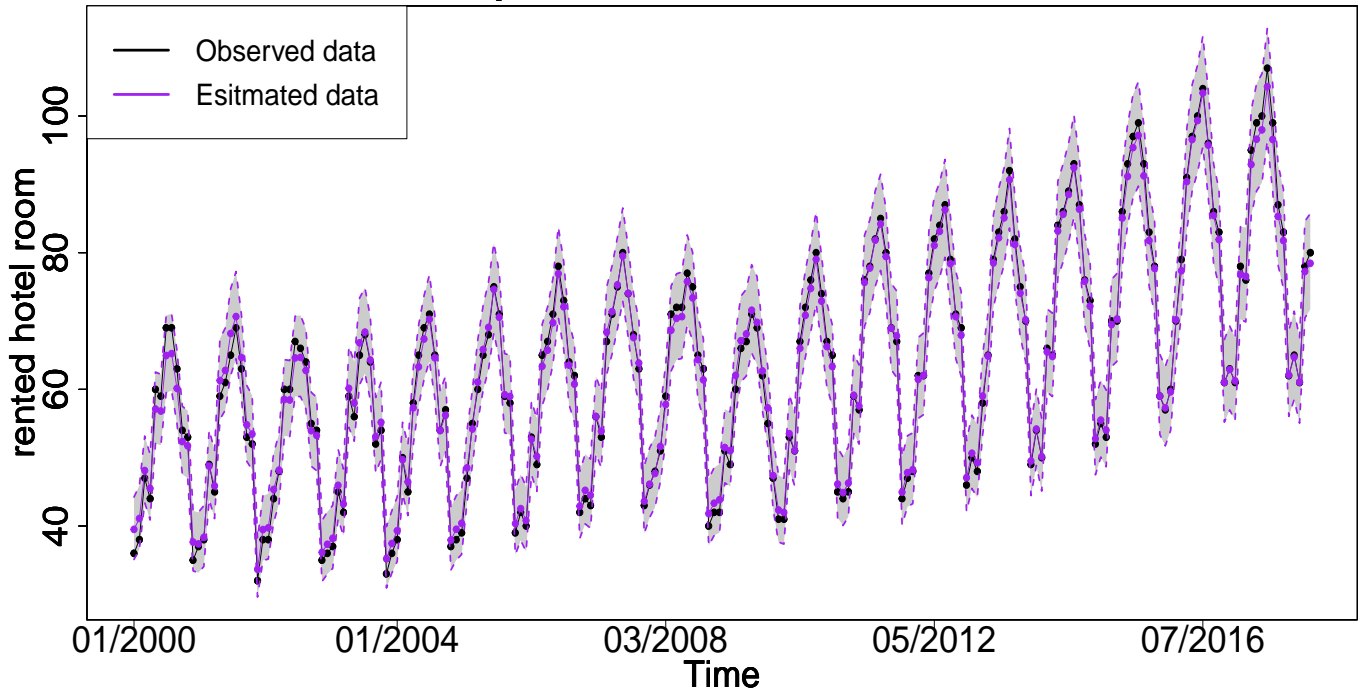
In-sample fit results of GLRGARMA model

Figure 5: In-sample fitting plot for Model 1 with $X = \text{PG}$

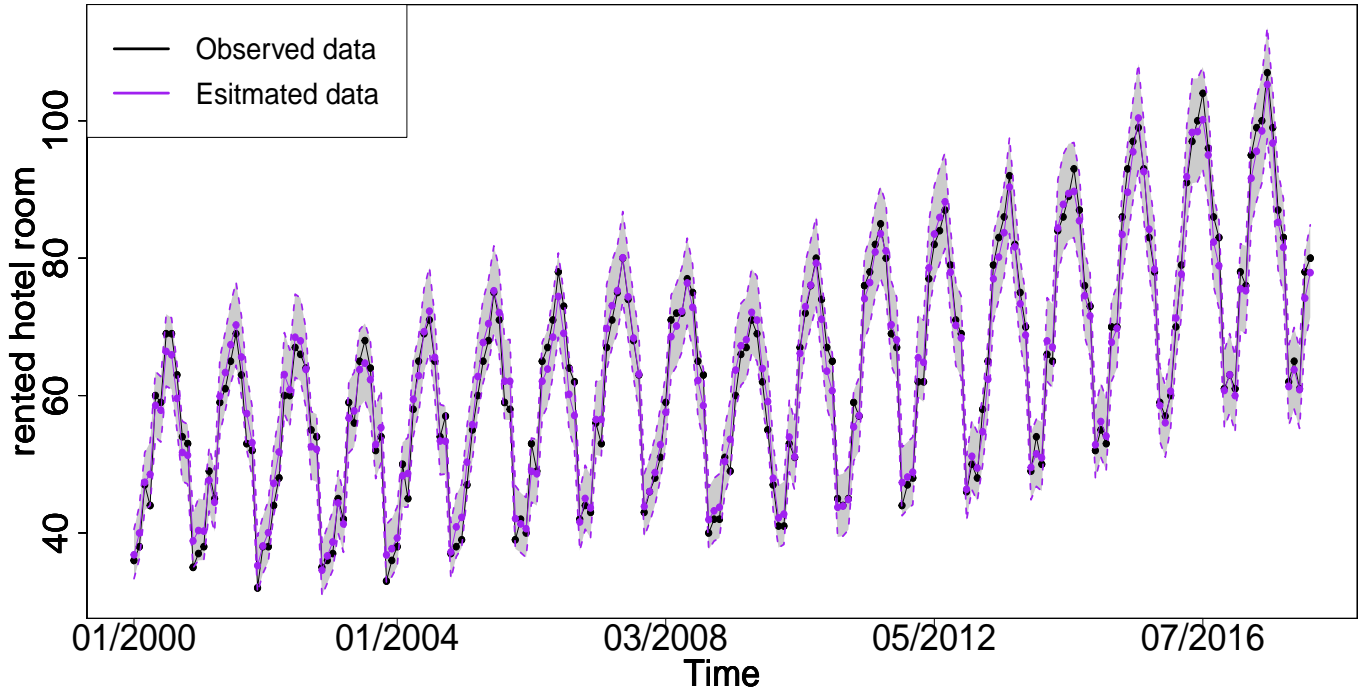
In-sample fit results of GLRGARMA model

Figure 6: In-sample fitting plot for Model 1 with $X = \text{UR}$

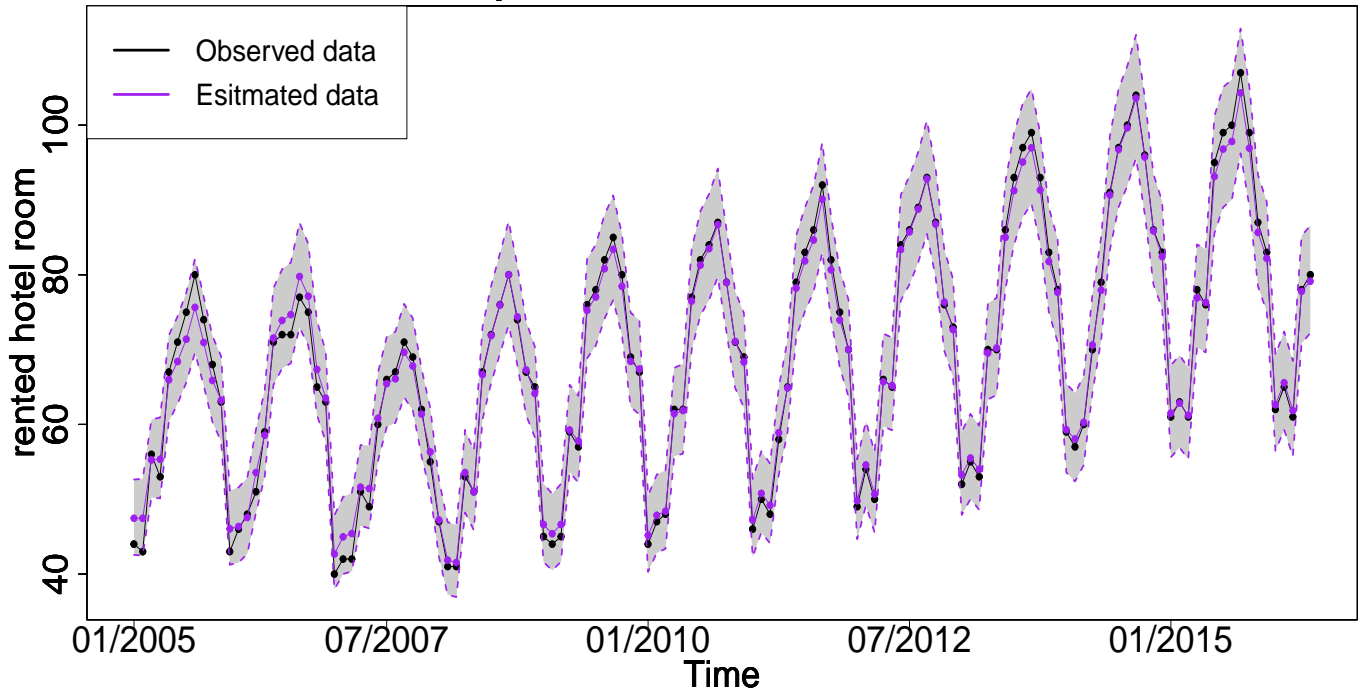
In-sample fit results of GLRSARMA model

Figure 7: In-sample fitting plot for Model 5 with $X = \text{IPI}$

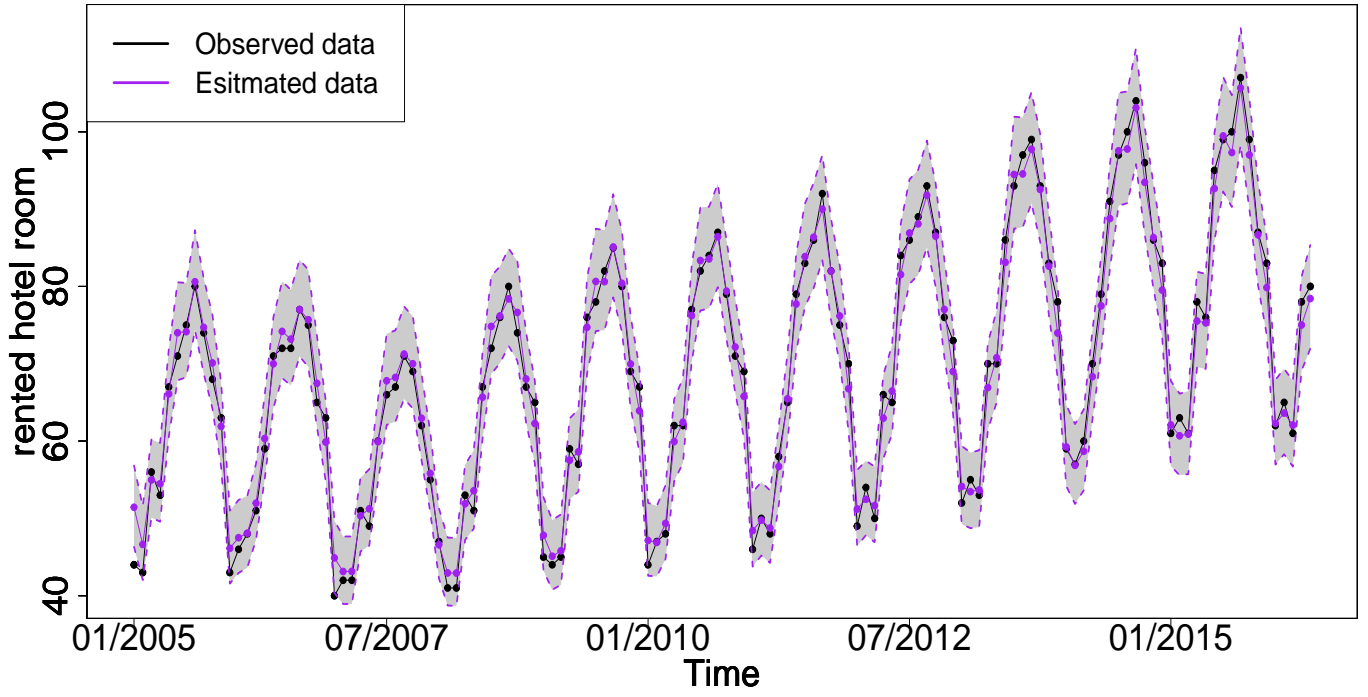
In-sample fit results of GLRSARMA model

Figure 8: In-sample fitting plot for Model 5 with $X = PG$

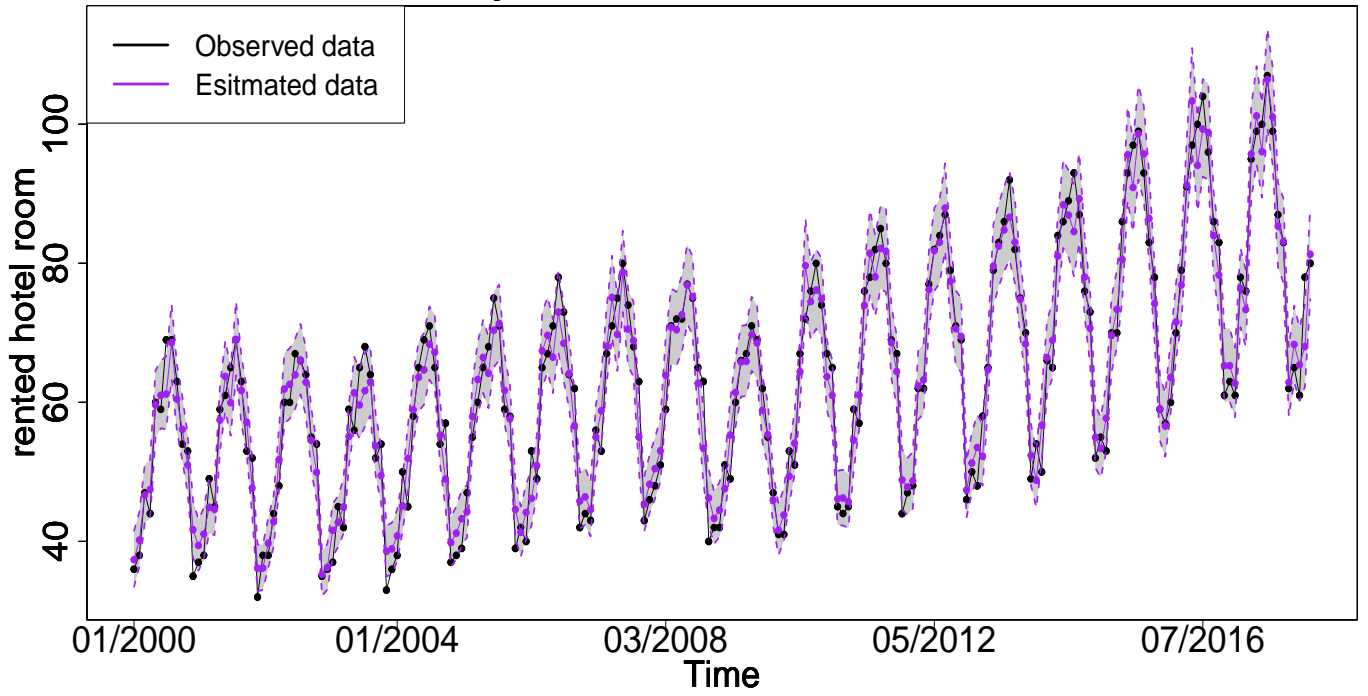
In-sample fit results of GLRSARMA model

Figure 9: In-sample fitting plot for Model 5 with $X = UR$

In-sample fit results of GLRGARMA model

Figure 10: In-sample fitting plot for Model 2 with $X_1 = \text{IPI}$ and $X_2 = \text{UR}$

In-sample fit results of GLRGARMA model

Figure 11: In-sample fitting plot for Model 2 with $X_1 = \text{IPI}$ and $X_2 = \text{PG}$

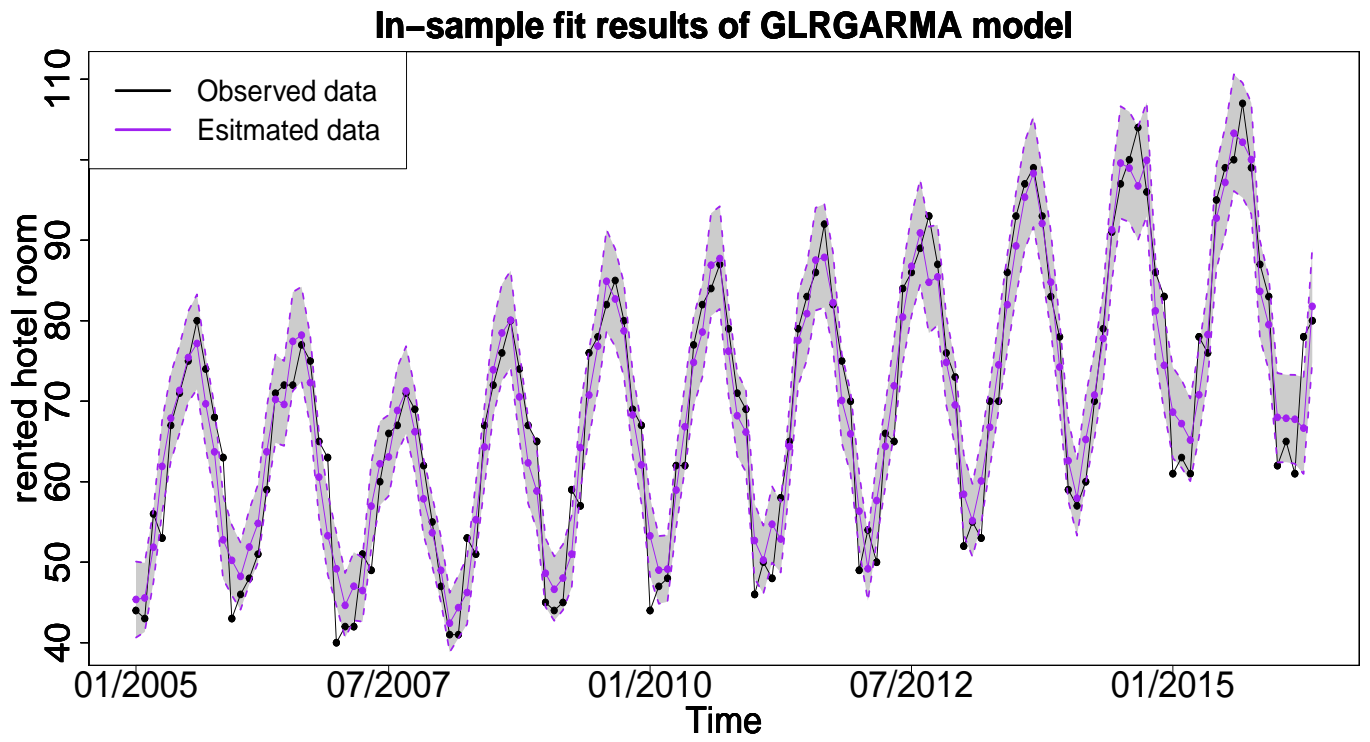


Figure 12: In-sample fitting plot for Model 2 with $X_1 = \text{PG}$ and $X_2 = \text{UR}$

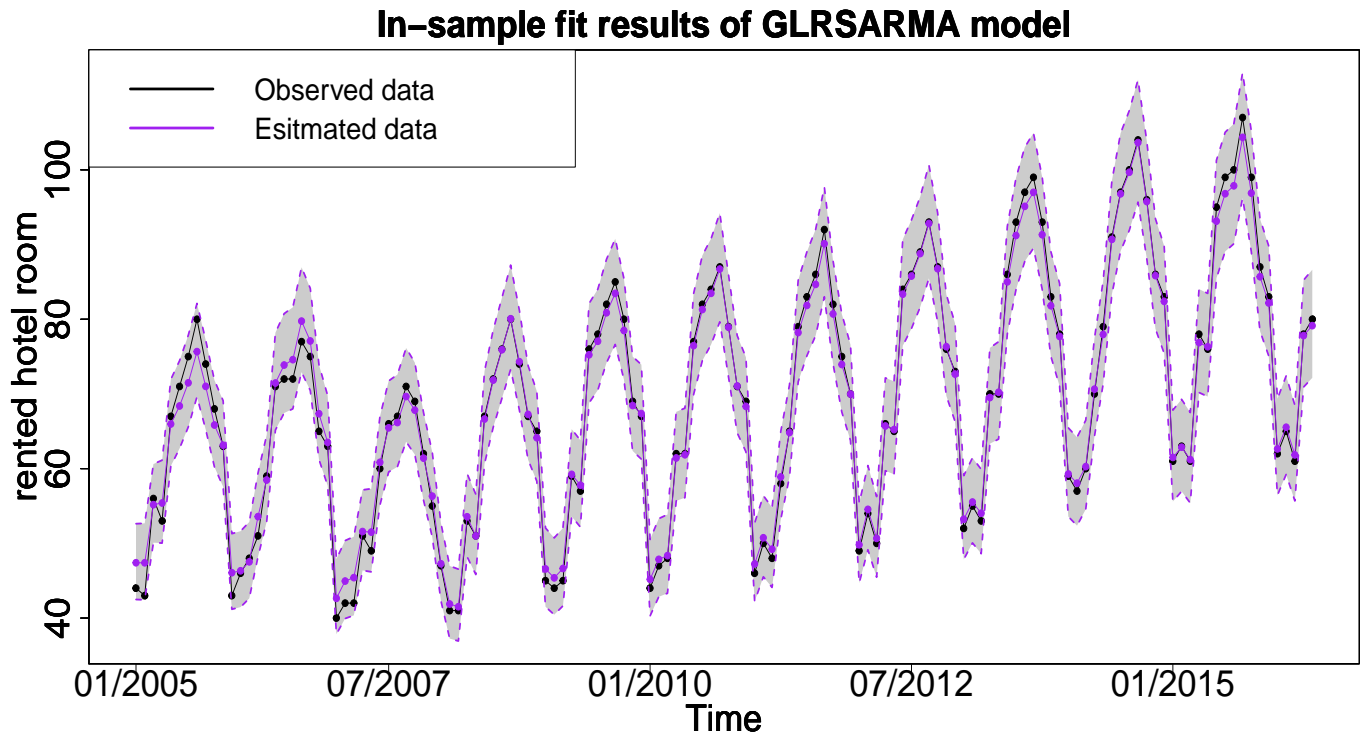


Figure 13: In-sample fitting plot for Model 6 with $X_1 = \text{IPI}$ and $X_2 = \text{UR}$

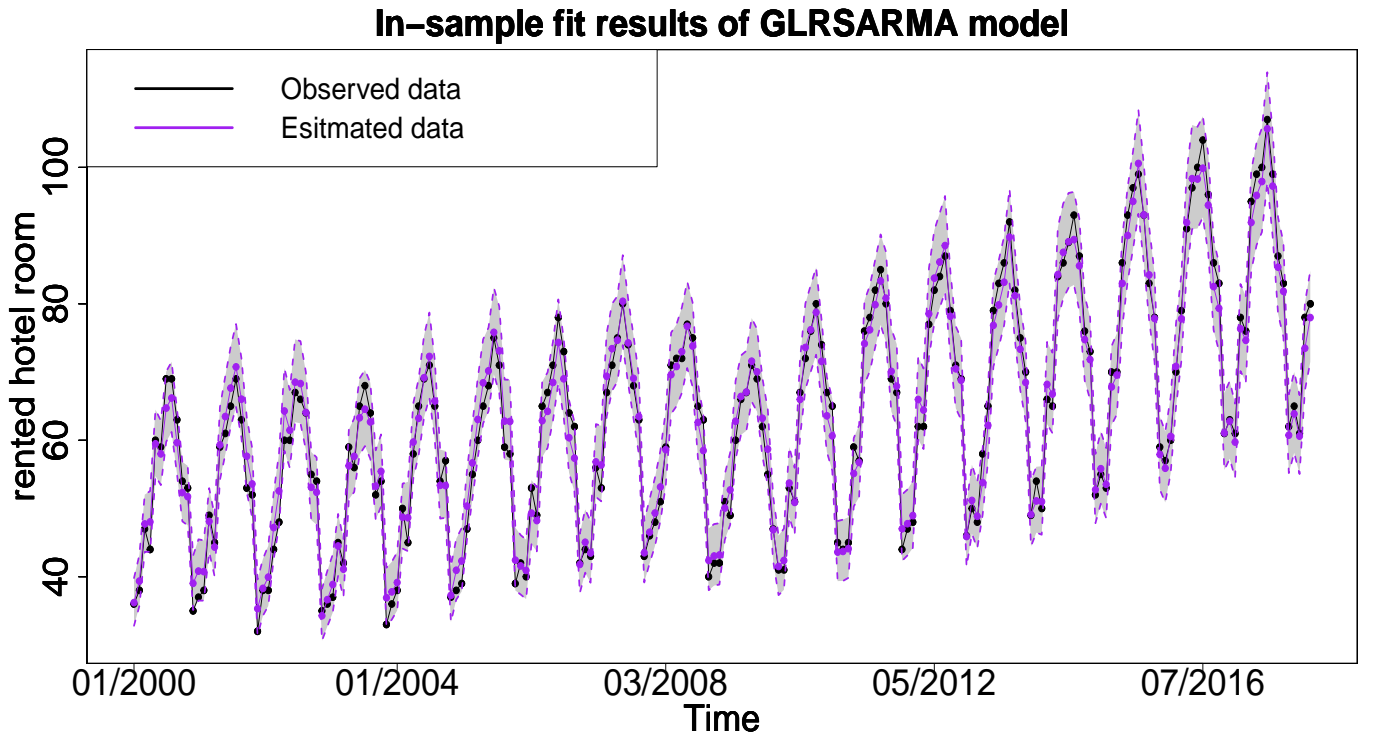


Figure 14: In-sample fitting plot for Model 6 with $X_1 = \text{IPI}$ and $X_2 = \text{PG}$

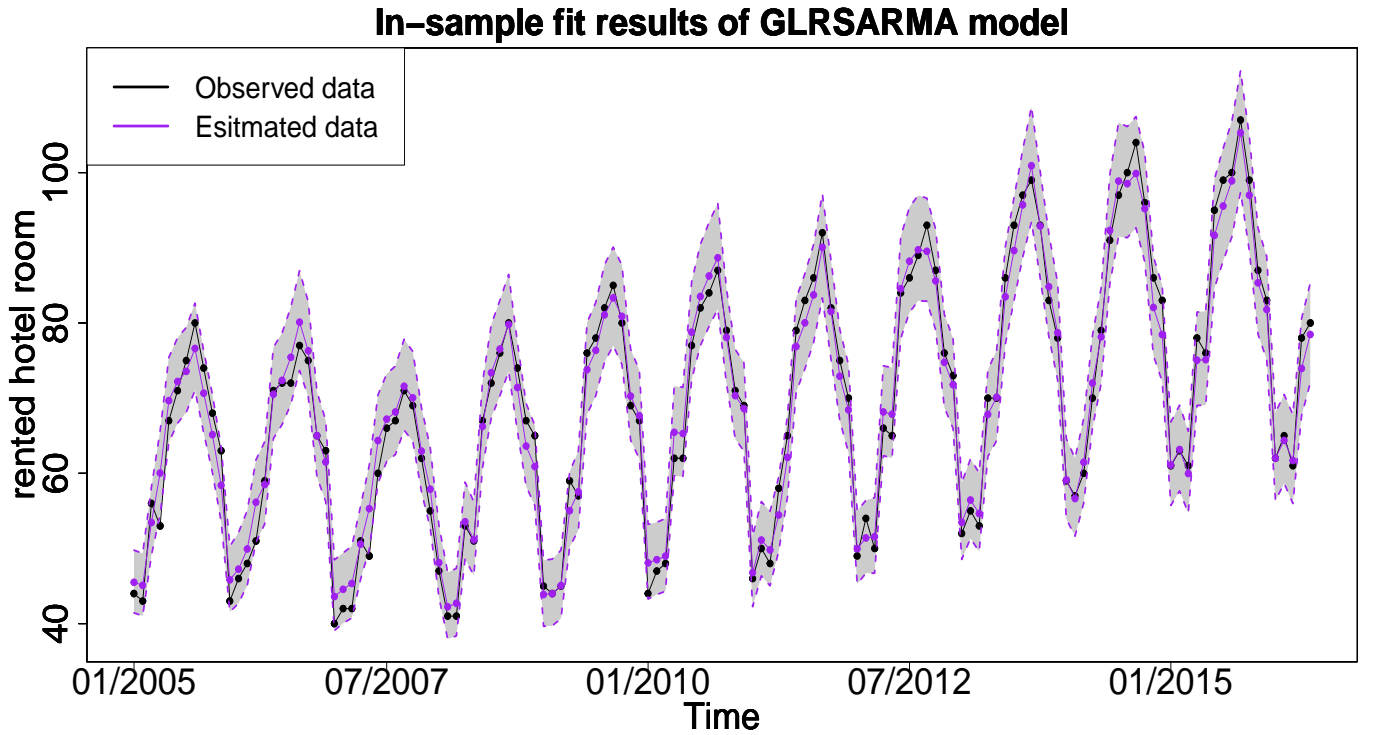


Figure 15: In-sample fitting plot for Model 6 with $X_1 = \text{PG}$ and $X_2 = \text{UR}$

4.3 Model forecast

In this section, we calculate one-step ahead forecast for $m = 20$ time points based on the posterior predictive distributions and the posterior sample size of $L = 90,000$. Only Model 1, Model 2, Model 5 and

Model 6 are adopted in out-sample forecasting study because these models show reasonable in-sample fitting performance. The forecasts \hat{y}_t are given by the posterior mean or median. To evaluate the forecast performance, three types of measures, namely residuals $r_t = y_t - \hat{y}_t$, percentage errors $p_t = \frac{r_t}{y_t} \times 100$ and scaled errors q_t are adopted. Based on r_t and p_t , three popular criteria, namely the mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) are calculated in Table 3. Overall, Gegenbauer long memory models with smaller criteria values provide more accurate forecast results than seasonal models. Moreover, the forecast performance can be greatly improved by incorporating more explanatory variables.

Table 3: Comparison of models in forecasts.

Model 1	$X = \text{IPI}$	$X = \text{UR}$	$X = \text{PG}$
MAE	15.43	10.71	8.08
RMSE	16.75	12.01	9.59
MAPE	0.16	0.11	0.09
MASE	1.72	1.34	0.90
Model 5	$X = \text{IPI}$	$X = \text{UR}$	$X = \text{PG}$
MAE	18.08	12.01	12.91
RMSE	20.57	13.16	14.89
MAPE	0.18	0.12	0.13
MASE	2.02	1.19	1.44
Model 2	$X_1 = \text{IPI}, X_2 = \text{PG}$	$X_1 = \text{IPI}, X_2 = \text{UR}$	$X_1 = \text{UR}, X_2 = \text{PG}$
MAE	6.91	9.91	7.63
RMSE	8.02	11.27	9.17
MAPE	0.07	0.10	0.08
MASE	0.77	1.10	0.85
Model 6	$X_1 = \text{IPI}, X_2 = \text{PG}$	$X_1 = \text{IPI}, X_2 = \text{UR}$	$X_1 = \text{UR}, X_2 = \text{PG}$
MAE	10.92	9.92	8.27
RMSE	12.77	11.54	9.84
MAPE	0.11	0.10	0.09
MASE	1.22	1.11	0.92

where HR is short for the number of rented hotel room, PG is short for power production, and UR is short for unemployment rate.

Figures below show the forecast results, which agree with Table 3. Gegenbauer long memory type model should be the best choice in dealing with the number of rented hotel room data with Gegenbauer long memory features. Seasonal models cannot replace Gegenbauer long memory model since some fundamental features cannot be capture by seasonal component.

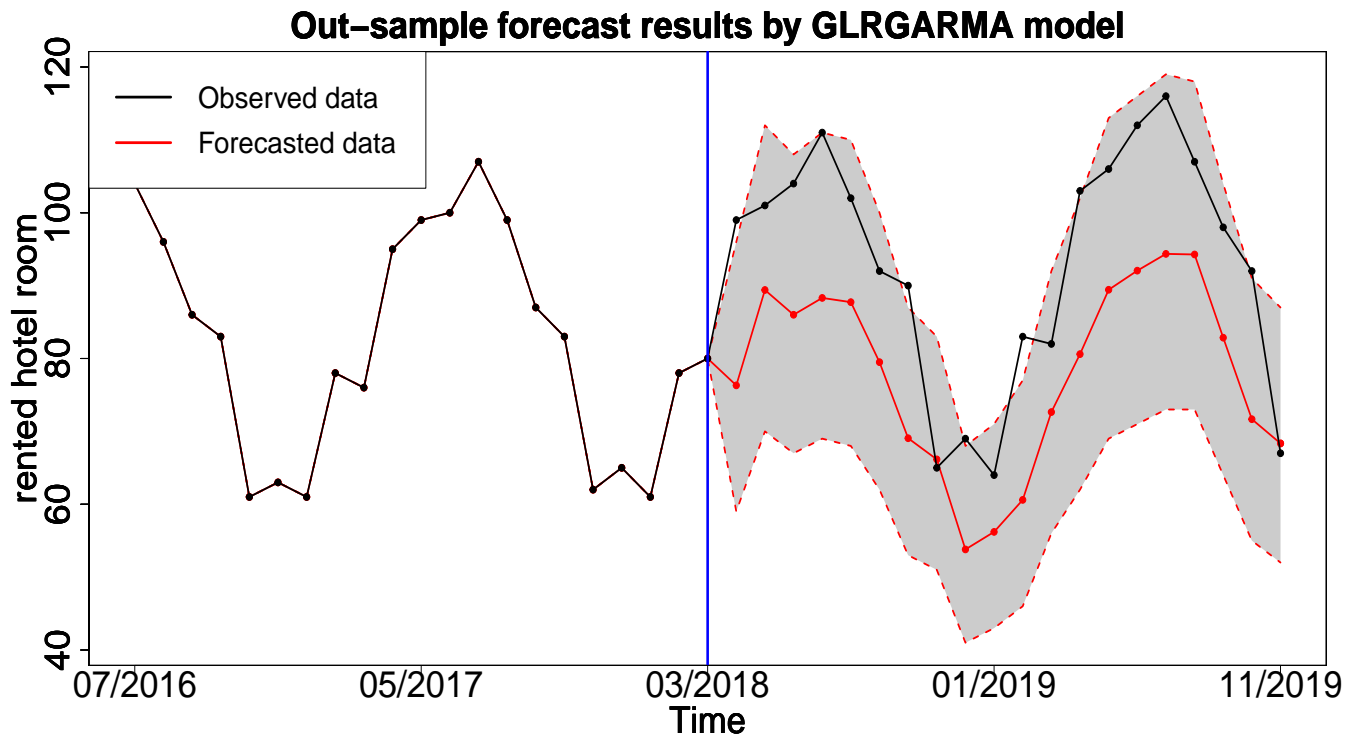


Figure 16: Out-sample forecasting plot for Model 1 with $X = \text{IPI}$

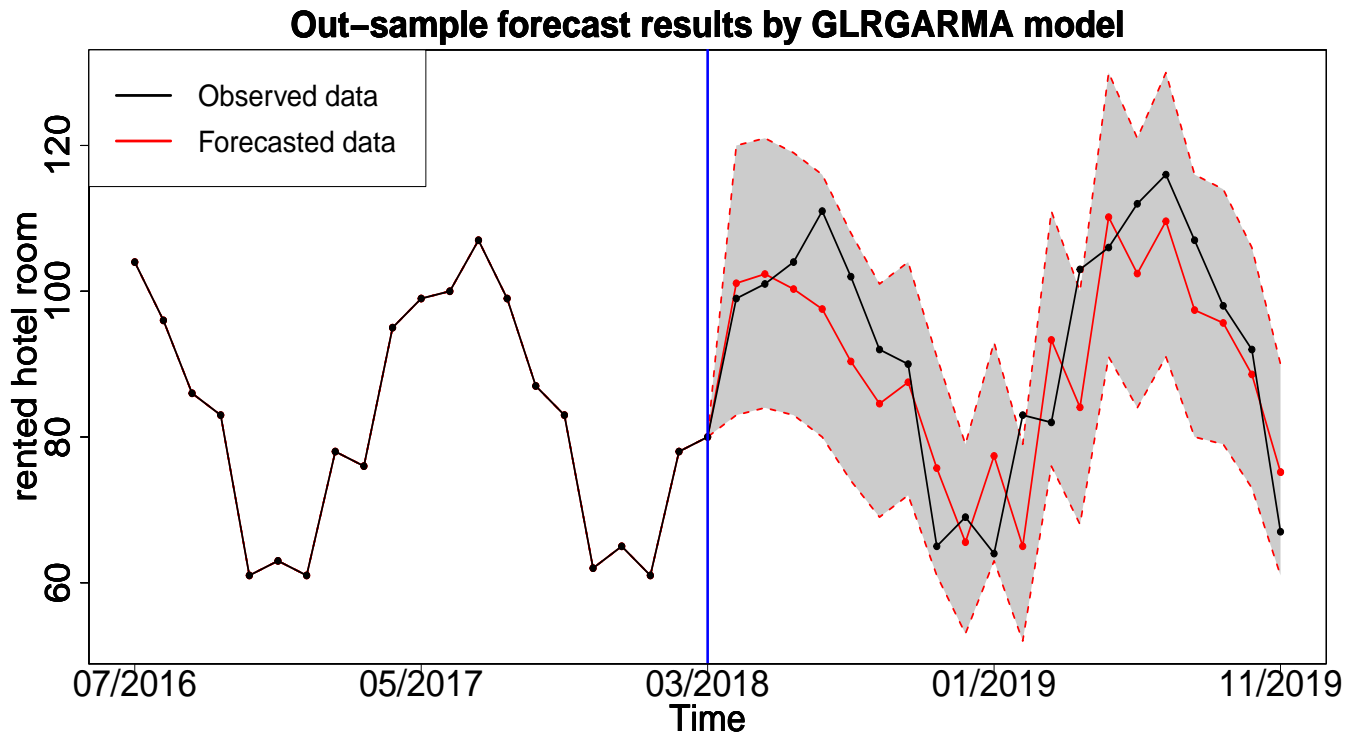


Figure 17: Out-sample forecasting plot for Model 1 with $X = \text{PG}$

Out-sample forecast results by GLRGARMA model

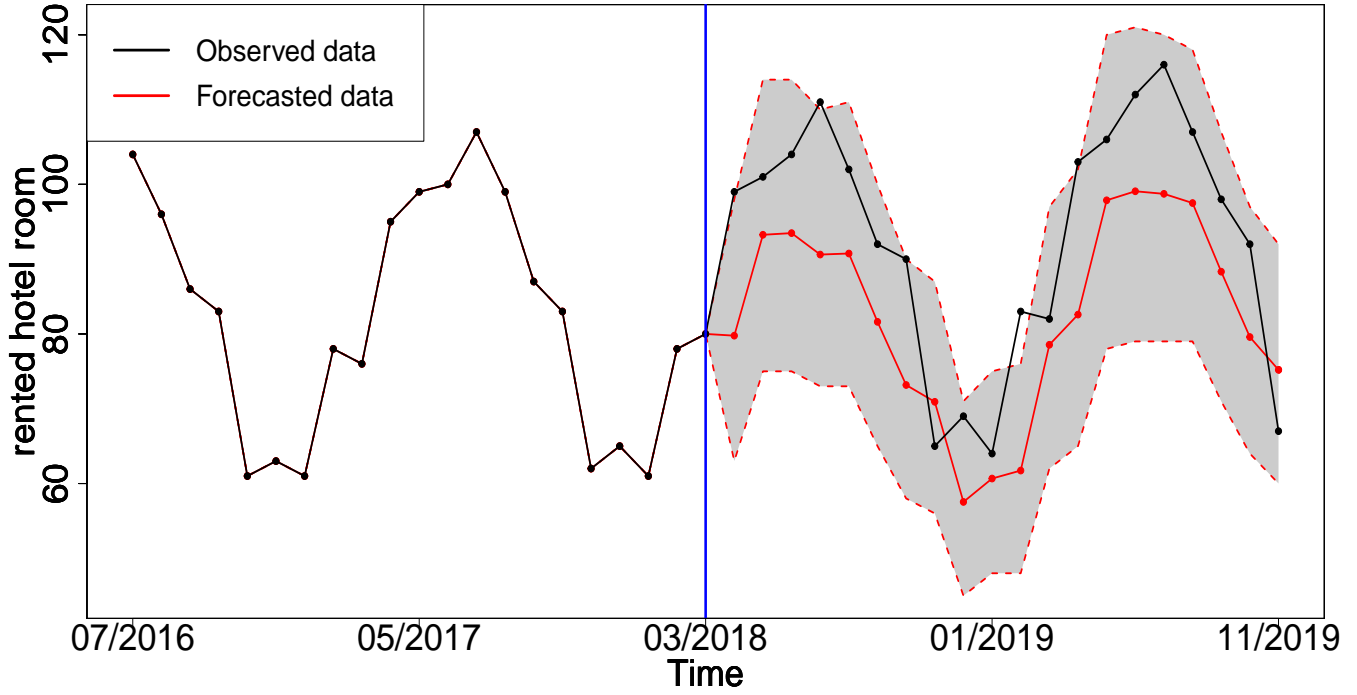


Figure 18: Out-sample forecasting plot for Model 5 with $X = \text{UR}$

Out-sample forecast results by GLRSARMA model

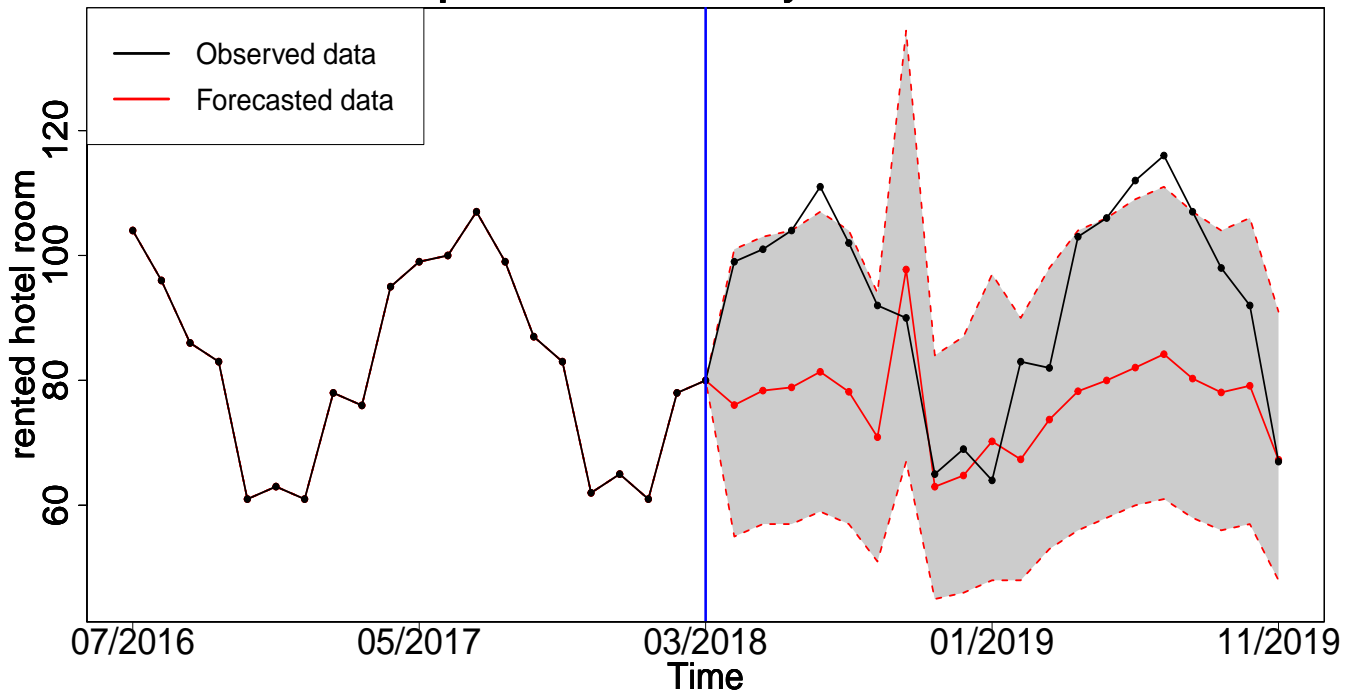


Figure 19: Out-sample forecasting plot for Model 5 with $X = \text{IPI}$

Out-sample forecast results by GLRSARMA model

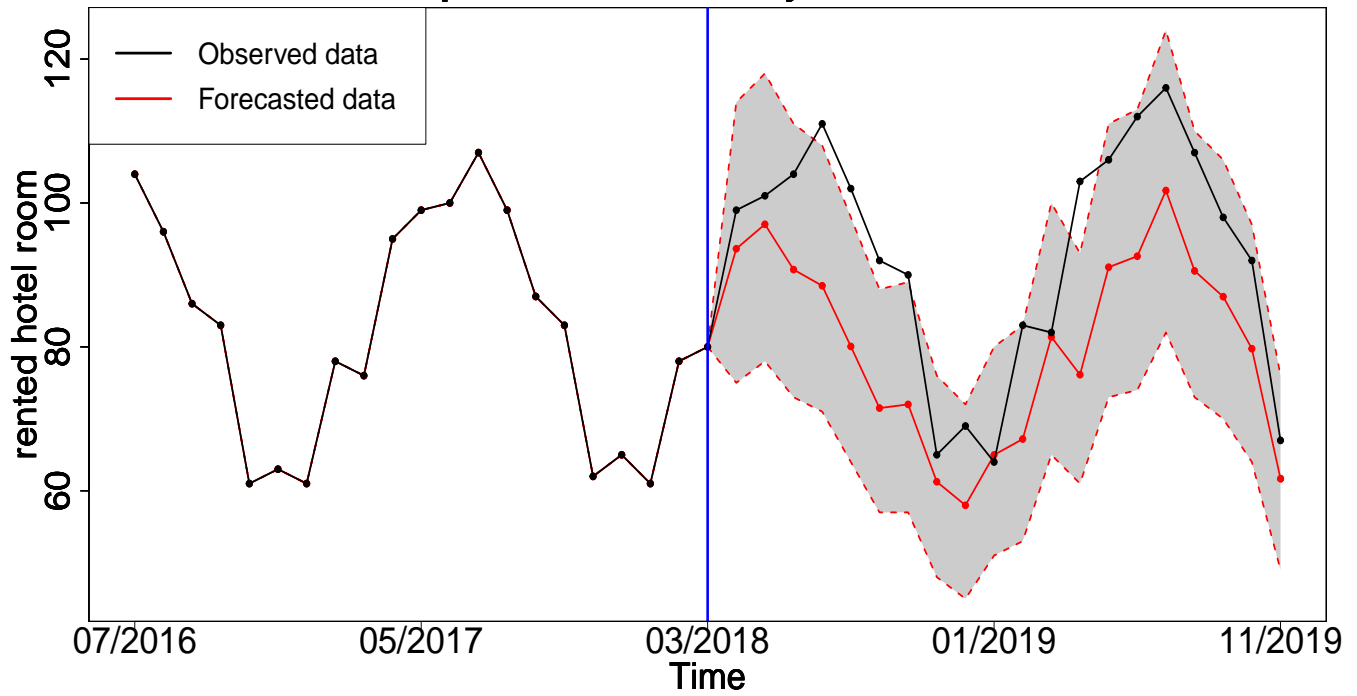


Figure 20: Out-sample forecasting plot for Model 5 with $X = PG$

Out-sample forecast results by GLRSARMA model

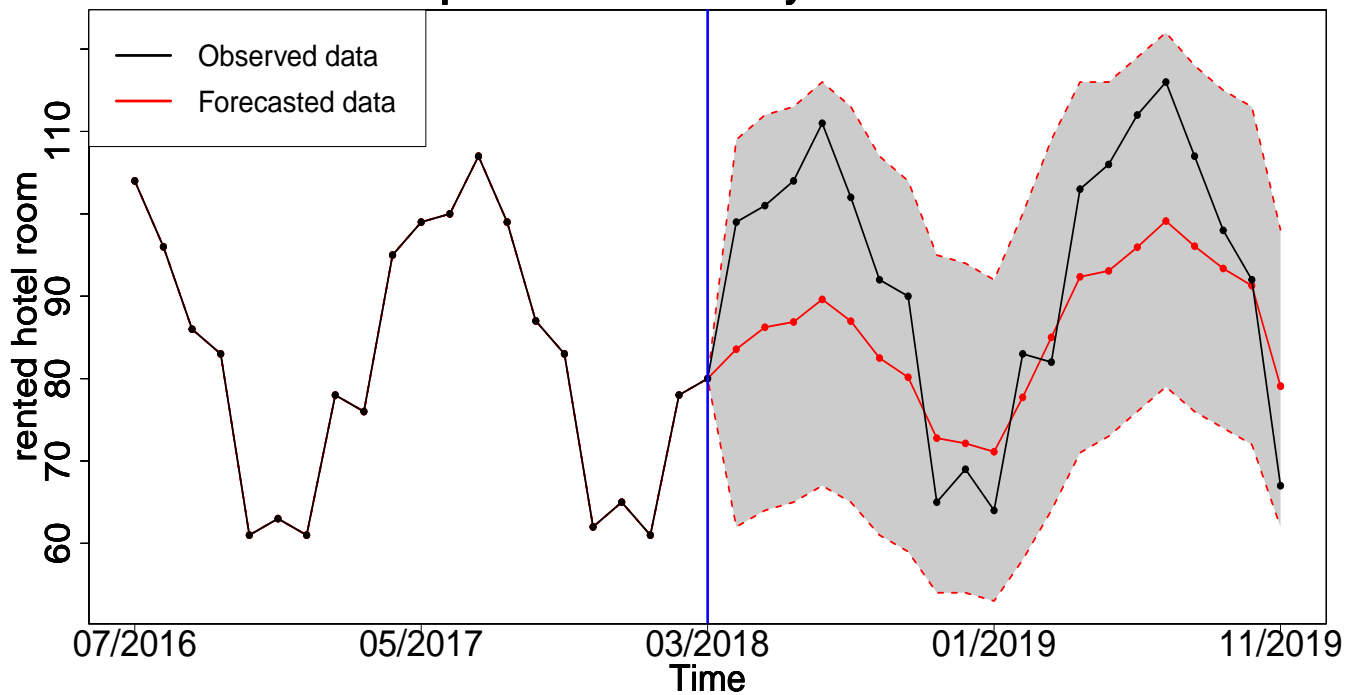


Figure 21: Out-sample forecasting plot for Model 5 with $X = UR$

Out-sample forecast results by GLRSARMA model

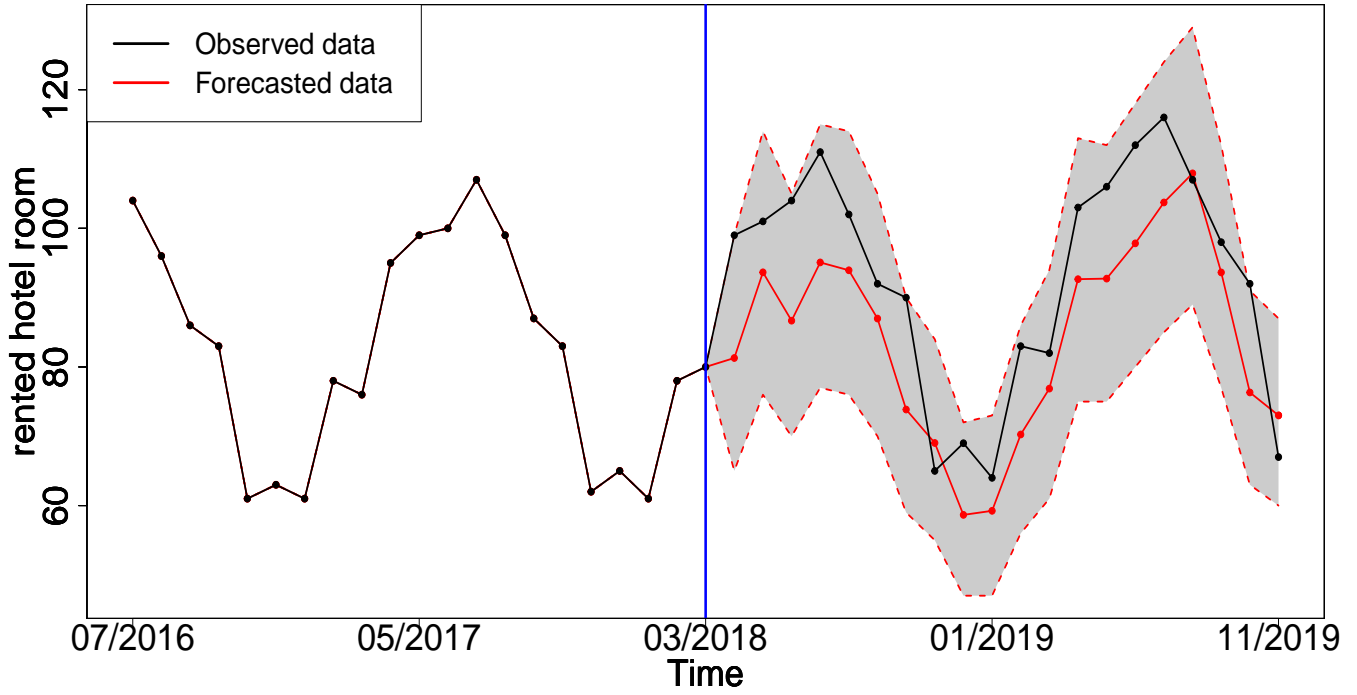


Figure 22: Out-sample forecasting plot for Model 2 with $X_1 = \text{IPI}$ and $X_2 = \text{UR}$

Out-sample forecast results by GLRSARMA model

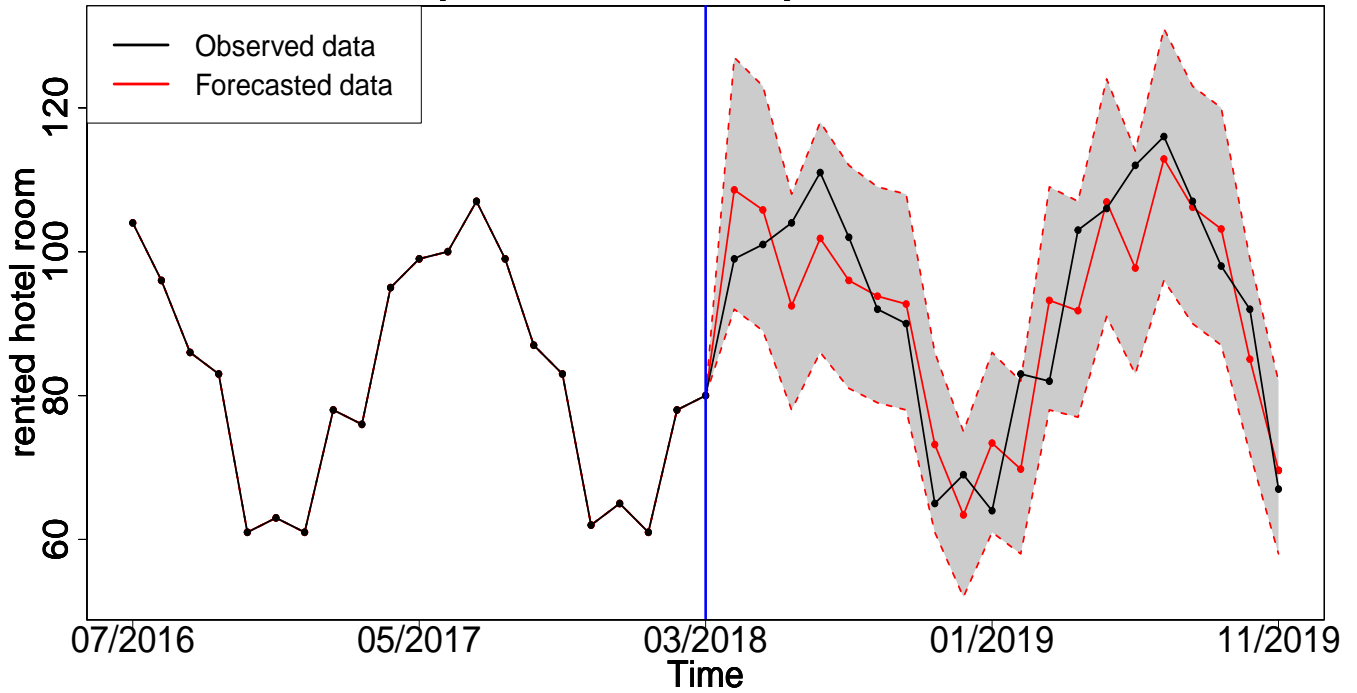


Figure 23: Out-sample forecasting plot for Model 2 with $X_1 = \text{IPI}$ and $X_2 = \text{PG}$

Out-sample forecast results by GLRSARMA model

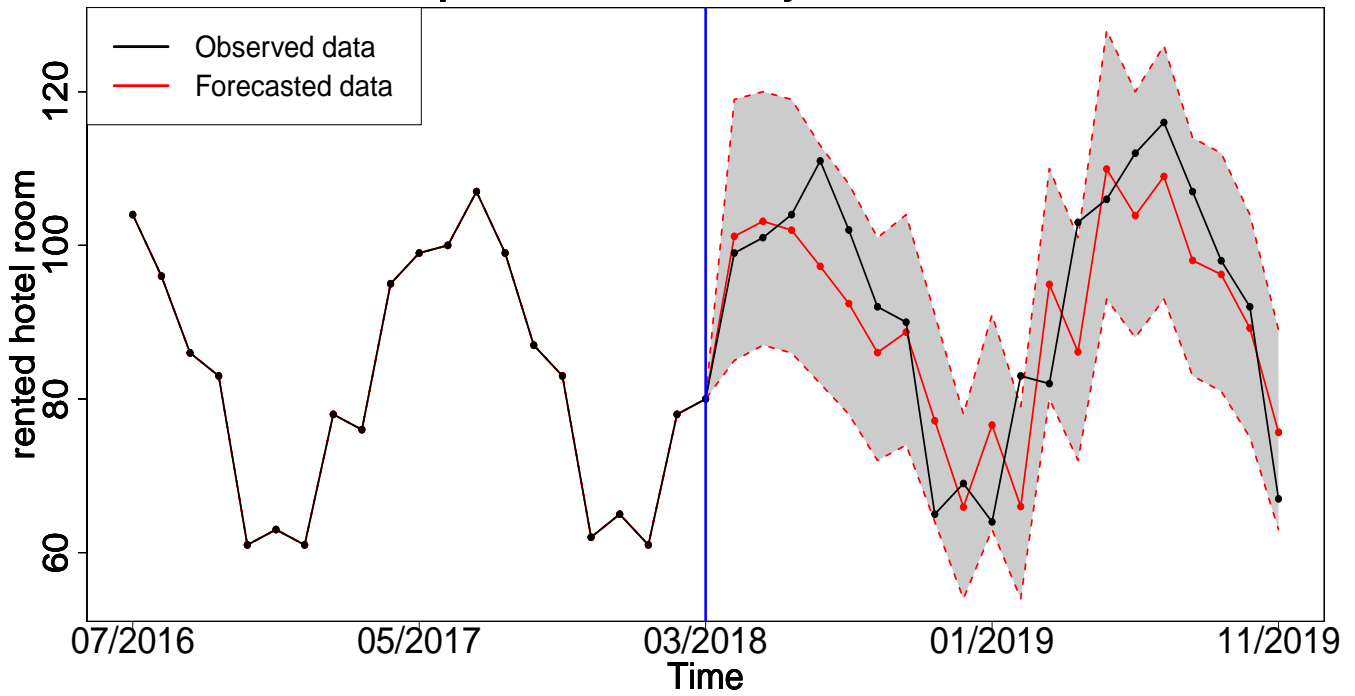


Figure 24: Out-sample forecasting plot for Model 2 with $X_1 = \text{PG}$ and $X_2 = \text{UR}$

Out-sample forecast results by GLRSARMA model

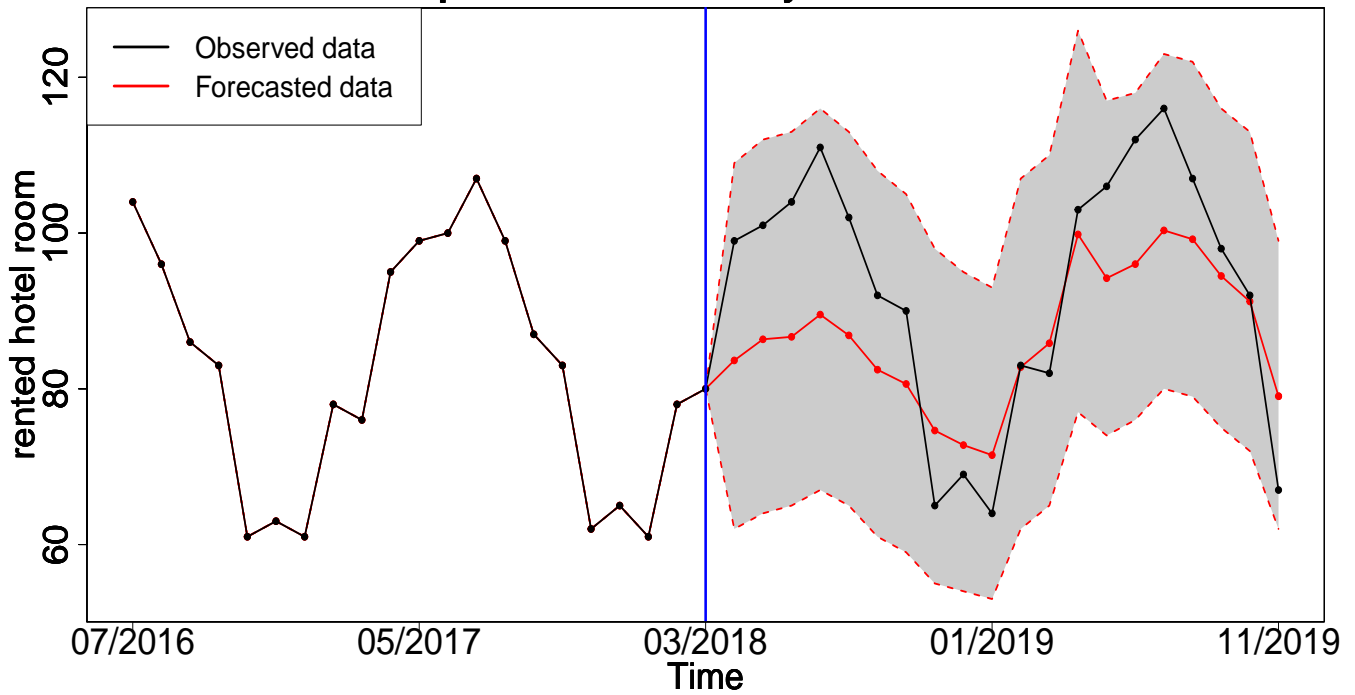


Figure 25: Out-sample forecasting plot for Model 6 with $X_1 = \text{IPI}$ and $X_2 = \text{UR}$

Out-sample forecast results by GLRSARMA model

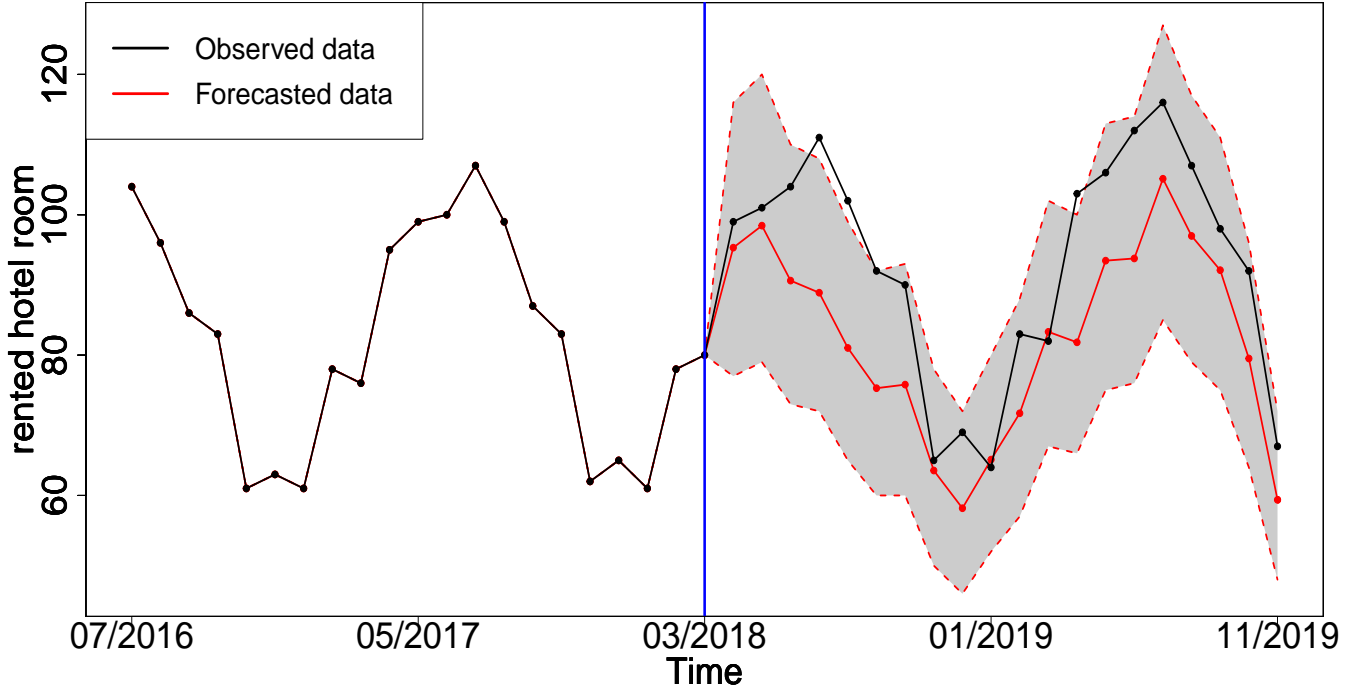


Figure 26: Out-sample forecasting plot for Model 6 with $X_1 = \text{IPI}$ and $X_2 = \text{PG}$

Out-sample forecast results by GLRSARMA model

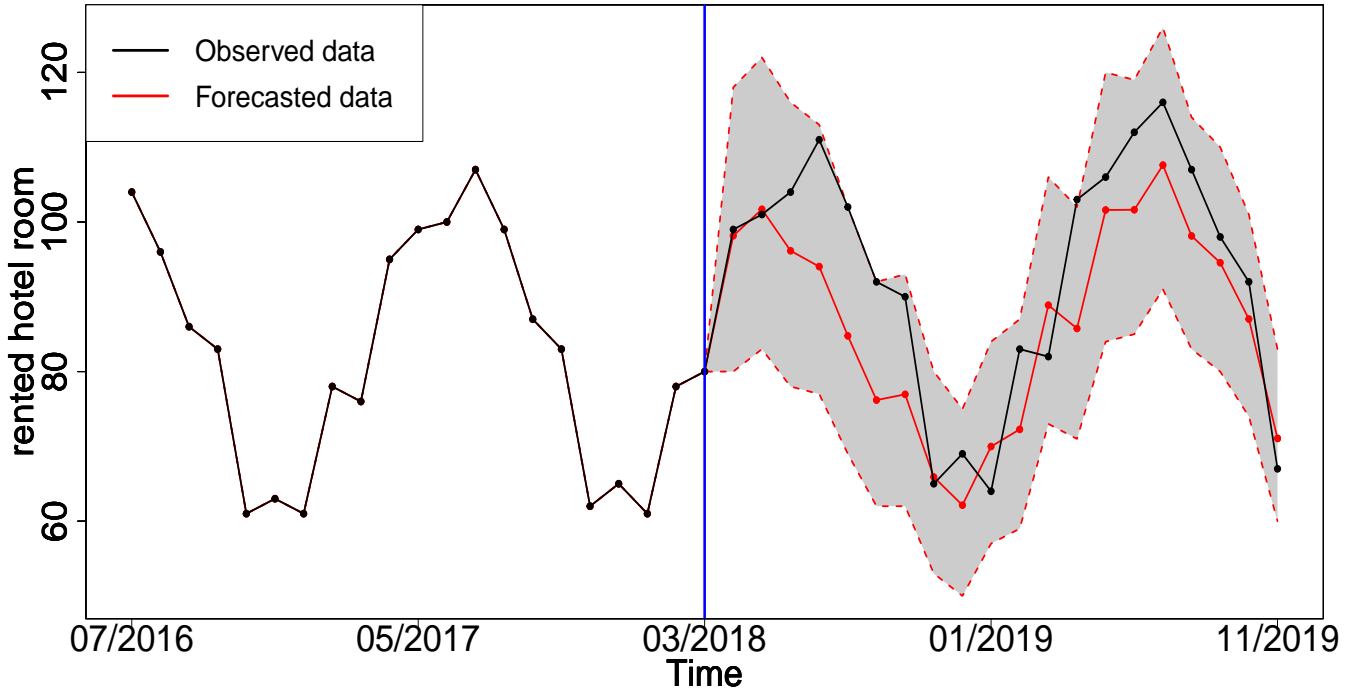


Figure 27: Out-sample forecasting plot for Model 6 with $X_1 = \text{PG}$ and $X_2 = \text{UR}$

5 Conclusion

This paper proposed a generalised linear regression structure with a innovative function of explanatory variables. Essential relevant information for modelling can be take into the consideration via explanatory

variables. To capture the Gegenbauer long memory feature in a time series, seasonal and Gegenbauer long memory components are incorporated to enhance model feasibility. Moreover, to improve model flexibility, the generalised Poisson (GP) distribution with over- equal- and under-dispersion is adopted. Overall, eight sub-models are implemented with the number of rented hotel room data set to evaluate the model performance. Furthermore, for the number of rented hotel room data in tourism area, the Gegenbauer long memory feature is revealed. Especially, the fundamental differences between seasonal component and Gegenbauer long memory component are distinguished. By plotting ACF and periodogram graphs, the long memory pattern is investigated. Bayesian approach is applied to implement in-sample fitting and out-sample forecast studies. Several model selection criteria are adopted to select most feasible model. Overall, GLRGARMA model is evaluated to be the best model to handle the time series with Gegenbauer long memory feature, especially in tourism area.

References

- (2008). World trade organization (wto). http://www.wto.org/english/res_e/statis_e/its2007_e/its07_merch_trade_product_e.htm.
- (2020). Statistics denmark. <https://www.dst.dk/en>.
- (2020). World bank country and lending groups. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519#High_income.
- Annis, A. and Lloyd, E. (1976). The expected value of the adjusted rescaled hurst range of independent normal summands. *Biometrika*, 63(1):111–116.
- Aynalem, S., Birhanu, K., and Tesefay, S. (2016). Employment opportunities and challenges in tourism and hospitality sectors. *Journal of Tourism & Hospitality*, 5(6):1–5.
- Beran, J. (1994). *Statistics for long-memory processes*, volume 61. CRC press.
- Bunghez, C. L. (2016). The importance of tourism to a destination’s economy. *Journal of Eastern Europe Research in Business & Economics*, 2016:1–9.
- Butler, R. (1998). Seasonality in tourism: Issues and implications. *The Tourist Review*.
- Cannas, R. (2012). An overview of tourism seasonality: key concepts and policies. *Almatourism-Journal of Tourism, Culture and Territorial Development*, 3(5):40–58.
- Chang, Y.-W. and Liao, M.-Y. (2010). A seasonal arima model of tourism forecasting: The case of taiwan. *Asia Pacific journal of Tourism research*, 15(2):215–221.
- Chen, J. L., Li, G., Wu, D. C., and Shen, S. (2019). Forecasting seasonal tourism demand using a multiseries structural time series method. *Journal of Travel Research*, 58(1):92–103.
- Chu, F.-L. (2009). Forecasting tourism demand with arma-based methods. *Tourism Management*, 30(5):740–751.
- Copeland, B. R. (1991). Tourism, welfare and de-industrialization in a small open economy. *Economica*, pages 515–529.
- Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

- Davis, R. A., Dunsmuir, W. T., and Wang, Y. (1999). Modeling time series of count data. *Statistics Textbooks and Monographs*, 158:63–114.
- Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, 195(2):216–222.
- Duro, J. A. and Turrión-Prats, J. (2019). Tourism seasonality worldwide. *Tourism Management Perspectives*, 31:38–53.
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81(395):709–721.
- Ferrante, M., Magno, G. L. L., and De Cantis, S. (2018). Measuring tourism seasonality across european countries. *Tourism Management*, 68:220–235.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.
- Gil-Alana, L. A., Cunado, J., and Perez de Gracia, F. (2008). Tourism in the canary islands: forecasting using several seasonal time series models. *Journal of Forecasting*, 27(7):621–636.
- Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.
- Graves, T., Gramacy, R. B., Watkins, N., and Franzke, C. (2014). A brief history of long memory. *arXiv preprint arXiv:1406.6018*.
- Groß-KlußMann, A. and Hautsch, N. (2013). Predicting bid–ask spreads using long-memory autoregressive conditional poisson models. *Journal of Forecasting*, 32(8):724–742.
- Hassler, U. (1994). (mis) specification of long memory in seasonal time series. *Journal of Time Series Analysis*, 15(1):19–30.
- Hosking, J. R. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–808.
- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.
- Jang, S. S. (2004). Mitigating tourism seasonality: A quantitative approach. *Annals of Tourism Research*, 31(4):819–836.
- Jolliffe, L. and Farnsworth, R. (2003). Seasonality in tourism employment: human resource challenges. *International Journal of Contemporary Hospitality Management*.
- Kenell, L. (2008). Dutch disease and tourism-the case of thailand.
- Koopmans, L. H. (1995). *The spectral analysis of time series*. Academic press.
- Kulendran, N. and Wong, K. K. (2005). Modeling seasonality in tourism forecasting. *Journal of Travel Research*, 44(2):163–170.

- Lew, A. A. (2011). Tourism's role in the global economy. *Tourism Geographies*, 13(1):148–151.
- Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Probability Theory and Related Fields*, 31(4):271–285.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
- Millen, S. and Beard, R. (2003). Estimation of the Hurst exponent for the Burdekin River using the Hurst-Mandelbrot rescaled range statistic. In *First Queensland Statistics Conference*.
- Mishra, P. K., Rout, H. B., and Pradhan, B. (2018). Seasonality in tourism and forecasting foreign tourist arrivals in india. *Iranian Journal of Management Studies*, 11(4):629–658.
- Neal, R. M. (1994). An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, 111(1):194–203.
- Neal, R. M. (2011). Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2:113–162.
- Ojaghlu, M. et al. (2019). Tourism-led growth and risk of the dutch disease: Dutch disease in turkey. *International Business Research*, 12(7):103–120.
- Peters, E. E. (1994). *Fractal market analysis: applying chaos theory to investment and economics*, volume 24. John Wiley & Sons.
- Porter-Hudak, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of the American Statistical Association*, 85(410):338–344.
- Qian, B. and Rasheed, K. (2004). Hurst exponent and financial market predictability. In *Proceedings of The 2nd IASTED international conference on financial engineering and applications*, pages 203–209.
- Rainville, E. D. (1960). *Special functions*, volume 442. New York.
- Refslund, B. and Sørensen, O. H. (2016). Islands in the stream? the challenges and resilience of the danish industrial relations model in a liberalising world. *Industrial Relations Journal*, 47(5-6):530–546.
- Saayman, A. and Botha, I. (2017). Non-linear models for tourism demand forecasting. *Tourism Economics*, 23(3):594–613.
- Shen, S., Li, G., and Song, H. (2009). Effect of seasonality treatment on the forecasting performance of tourism demand models. *Tourism Economics*, 15(4):693–708.
- Shen, S., Li, G., and Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research*, 38(1):72–89.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.
- Stein, E. M. and Weiss, G. L. (1971). *Introduction to Fourier analysis on Euclidean spaces*, volume 1. Princeton university press.

- Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and its Applications*, 312(1):285–299.
- Wold, H. (1938). *A study in the analysis of stationary time series*. PhD thesis, Almqvist & Wiksell.
- Woodward, W. A., Cheng, Q. C., and Gray, H. L. (1998). A k-factor GARMA long-memory model. *Journal of Time Series Analysis*, 19(4):485–504.
- Yan, H., Chan, J. S., and Peters, G. W. (2017). Long memory models for financial time series of counts and evidence of systematic market participant trading behaviour patterns in futures on US treasuries. *SSRN*: <https://ssrn.com/abstract=2962341>.